

Universal recovery from a decrease of quantum relative entropy

Marius Junge¹, Renato Renner², David Sutter², Mark M. Wilde³, and Andreas Winter⁴

¹*Department of Mathematics, University of Illinois at Urbana-Champaign, Illinois 61801-2975, USA*

²*Institute for Theoretical Physics, ETH Zurich, Switzerland*

³*Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA*

⁴*ICREA & Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain.*

Abstract

The data processing inequality states that the quantum relative entropy between two states ρ and σ can never increase by applying the same quantum channel \mathcal{N} to both states. This inequality can be strengthened with a remainder term in the form of a distance between ρ and the closest recovered state $(\mathcal{R} \circ \mathcal{N})(\rho)$, where \mathcal{R} is a recovery map with the property that $\sigma = (\mathcal{R} \circ \mathcal{N})(\sigma)$. We show the existence of an *explicit* recovery map that is *universal* in the sense that it depends only on σ and the quantum channel \mathcal{N} to be reversed. This result gives an alternate, information-theoretic characterization of the conditions for approximate quantum error correction.

1 Introduction

Suppose that an experiment, depending on an unknown parameter $\theta \in \Theta$, is described by classical probability distributions ν_θ on a sample space (X, Ω) . A *statistic*, i.e., a measurable map $\mathcal{N} : (X, \Omega) \rightarrow (Y, \Sigma)$ is called *sufficient* with respect to Θ [12], if the conditional probability does not depend on θ , i.e., if there exists a probability distribution ν on (X, Ω) such that

$$\nu_\theta(X|\mathcal{N}(X)) = \nu(X|\mathcal{N}(X)) \quad \text{for all } \theta \in \Theta. \quad (1)$$

In such a case, $\mathcal{N}(X)$ contains the same information about the parameter $\theta \in \Theta$ as X does.

This concept has been generalized to the quantum setup [34, 35, 31, 21]. For two Hilbert spaces A and B , let $S(A)$ denote the set of density operators on A and let $\text{TPCP}(A, B)$ be the set of trace-preserving completely positive maps from A to B . Let $Q(A)$ denote some subset of $S(A)$. A quantum channel $\mathcal{N} \in \text{TPCP}(A, B)$ is called *sufficient* (or *reversible*) *with respect to* $Q(A)$, if there exists a recovery map $\mathcal{R} \in \text{TPCP}(B, A)$ such that

$$(\mathcal{R} \circ \mathcal{N})(\rho) = \rho \quad \text{for all } \rho \in Q(A). \quad (2)$$

The *data processing inequality* (also known as *monotonicity of the relative entropy*) states that the relative entropy between two states ρ and σ — defined as $D(\rho\|\sigma) := \text{tr}(\rho(\log \rho - \log \sigma))$ if $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ and $+\infty$ otherwise — is non-increasing under physical evolutions [29, 44], i.e., $D(\rho\|\sigma) \geq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$, where \mathcal{N} is a quantum channel. This fundamental entropy inequality has many applications in quantum information theory and is closely related to the sufficiency of \mathcal{N} . As shown in [34, 35, 21, 22], for a given $\sigma \geq 0$ a quantum channel $\mathcal{N} \in \text{TPCP}(A, B)$ is sufficient with respect to $Q(A)$ if and

only if $D(\rho\|\sigma) = D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$ for all $\rho \in \mathcal{Q}(A)$. (We assume that the set $\mathcal{Q}(A)$ is such that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ for all $\rho \in \mathcal{Q}(A)$.) Furthermore, it is known that the data processing inequality holds with equality if and only if there exists a recovery map $\mathcal{R} \in \text{TPCP}(B, A)$ that simultaneously reverses the action of the physical evolution \mathcal{N} on both states [34, 35, 36], i.e., $(\mathcal{R} \circ \mathcal{N})(\rho) = \rho$ and $(\mathcal{R} \circ \mathcal{N})(\sigma) = \sigma$. If the Hilbert spaces A and B are assumed to be separable and $\sigma \in \text{TC}(A)$ is positive semi-definite, then the recovery map can be taken as the *Petz recovery map* $\mathcal{P}_{\sigma, \mathcal{N}}$ (also known as the *transpose map*), defined as the unique solution to

$$\langle a_2, \mathcal{N}^\dagger(a_1) \rangle_\sigma = \langle \mathcal{P}_{\sigma, \mathcal{N}}^\dagger(a_2), a_1 \rangle_{\mathcal{N}(\sigma)} \quad \text{for all } a_1, a_2 \in \text{L}(B), \quad (3)$$

where $\text{L}(B)$ denotes the set of bounded operators on B , $\text{TC}(A)$ the set of trace-class operators on A , $\langle a, b \rangle_\omega := \text{tr}(a^\dagger \omega^{\frac{1}{2}} b \omega^{\frac{1}{2}})$ is a weighted inner product with ω taken to be positive semi-definite and trace class [34, 35, 36], and \mathcal{N}^\dagger denotes the adjoint map of \mathcal{N} . The Petz recovery map is completely positive and trace non-increasing (see, e.g., [33]). In the case that the Hilbert spaces A and B are finite-dimensional, then, on the support of $\mathcal{N}(\sigma)$, the Petz recovery map takes the form

$$\mathcal{P}_{\sigma, \mathcal{N}} : X_B \mapsto \sigma^{\frac{1}{2}} \mathcal{N}^\dagger(\mathcal{N}(\sigma)^{-\frac{1}{2}} X_B \mathcal{N}(\sigma)^{-\frac{1}{2}}) \sigma^{\frac{1}{2}}. \quad (4)$$

(Following the standard convention, σ^{-1} is defined to be the inverse of σ , when σ is considered as an operator acting only on the support of σ .)

The concept of sufficient statistics can be made robust. A quantum channel $\mathcal{N} \in \text{TPCP}(A, B)$ is ε -*sufficient* with respect to $\mathcal{Q}(A)$ if there exists a recovery map $\mathcal{R}_\varepsilon \in \text{TPCP}(B, A)$ such that [20]

$$\frac{1}{2} \|\rho - (\mathcal{R}_\varepsilon \circ \mathcal{N})(\rho)\|_1 \leq \varepsilon \quad \text{for all } \rho \in \mathcal{Q}(A), \quad (5)$$

for some $\varepsilon \in [0, 1]$. Together with the case $\varepsilon = 0$ discussed above, this motivates the question if there exists a refined version of the data processing inequality in terms of recoverability [47], which would serve as an alternative characterization of approximate sufficient statistics.

An inequality that is closely related to the monotonicity of the relative entropy is the *strong subadditivity* of quantum entropy [26, 27], which ensures that for any tripartite state $\rho_{ABC} \in \mathcal{S}(A \otimes B \otimes C)$ the conditional mutual information is non-negative, i.e., $I(A : C|B)_\rho := H(AB)_\rho + H(BC)_\rho - H(ABC)_\rho - H(B)_\rho \geq 0$, where $H(A)_\rho := -\text{tr}(\rho_A \log \rho_A)$ denotes the von Neumann entropy. This inequality has been strengthened recently with a remainder term in the form of a distance to the closest recovered state. It was shown in [11], that for any density operator ρ_{ABC} there exists a trace-preserving completely positive map (the *recovery map*) $\mathcal{R}_{B \rightarrow BC}$ such that

$$I(A : C|B)_\rho \geq -2 \log F(\rho_{ABC}, \mathcal{R}_{B \rightarrow BC}(\rho_{AB})), \quad (6)$$

where the fidelity of ρ and σ is defined by $F(\rho, \sigma) := \|\sqrt{\rho} \sqrt{\sigma}\|_1$. If A , B , and C are finite-dimensional Hilbert spaces, on the support of ρ_B , $\mathcal{R}_{B \rightarrow BC}$ can be taken as a rotated Petz recovery map, i.e., a trace-preserving completely positive map of the form

$$X_B \mapsto V_{BC} \rho_{BC}^{\frac{1}{2}} (\rho_B^{-\frac{1}{2}} U_B X_B U_B^\dagger \otimes \text{id}_C) \rho_{BC}^{\frac{1}{2}} V_{BC}^\dagger, \quad (7)$$

where V_{BC} and U_B are unitaries on $B \otimes C$ and B , respectively.

The result of [11], whose proof is based on de Finetti type arguments and properties of Rényi entropies, has been extended and generalized in various ways. In [7], based on the quantum state redistribution protocol [9] and de Finetti type arguments, it was shown that the fidelity term can be replaced by a *measured relative entropy* $D_{\mathbb{M}}$, which is never smaller than the fidelity term, i.e.,

$$I(A : C|B)_\rho \geq D_{\mathbb{M}}(\rho_{ABC} \|\mathcal{R}_{B \rightarrow BC}(\rho_{AB})) \geq -2 \log F(\rho_{ABC}, \mathcal{R}_{B \rightarrow BC}(\rho_{AB})). \quad (8)$$

The measured relative entropy is defined as the supremum of the relative entropy with measured inputs over all positive operator-valued measures (POVMs) $\mathcal{M} = \{M_x\}$, i.e.,

$$D_{\mathbb{M}}(\rho \|\sigma) := \sup \left\{ D(\mathcal{M}(\rho) \|\mathcal{M}(\sigma)) : \mathcal{M}(\rho) = \sum_x \text{tr}(\rho M_x) |x\rangle\langle x| \text{ with } \sum_x M_x = \text{id} \right\}, \quad (9)$$

where $\{|x\rangle\}$ is a finite set of orthonormal vectors. We note that the tighter bound from [7] came at the cost of losing all information about the structure of the recovery map. In [42], it was shown that there exists a recovery map both satisfying (8) and possessing a universality property, in the sense that it only depends on the marginal ρ_{BC} . Furthermore, for a linearized version of (6) it was shown that the recovery map has the form of a rotated Petz recovery map with commuting unitaries, i.e., a recovery map of the form in (7) where V_{BC} and U_B commute with ρ_{BC} and ρ_B , respectively.

In view of approximate sufficiency of quantum channels discussed above, it would be helpful to have a generalization of (6) in terms of relative entropies. This has been established in [45] with a proof technique based on the notion of a Rényi generalization of a relative entropy difference [39] and Hadamard's three-line theorem. It was shown that for any two states ρ and σ with $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ and any channel \mathcal{N} there exists a recovery map \mathcal{R} such that $(\mathcal{R} \circ \mathcal{N})(\sigma) = \sigma$ and

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq -2 \log F(\rho, (\mathcal{N} \circ \mathcal{R})(\rho)) . \quad (10)$$

Furthermore, the recovery map was shown to be a rotated Petz recovery map with unitaries U and V in the algebra generated by σ and $\mathcal{N}(\sigma)$, respectively. Very recently, another different proof technique was found [43], based on the concavity and monotonicity of the operator logarithm, which shows that there exists a recovery map \mathcal{R} such that

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq D_{\mathbb{M}}(\rho\|(\mathcal{R} \circ \mathcal{N})(\rho)) \quad (11)$$

$$\geq -2 \log F(\rho, (\mathcal{N} \circ \mathcal{R})(\rho)) . \quad (12)$$

The recovery map was shown to be a convex combination of rotated Petz recovery maps with unitaries U and V in the algebra generated by σ and $\mathcal{N}(\sigma)$, respectively, and therefore satisfies $(\mathcal{R} \circ \mathcal{N})(\sigma) = \sigma$.

Neither in [45] nor in [43] could the recovery map satisfying (10) and (11), respectively, be shown to be universal, in the sense that it could be taken independent of ρ . In this article, we show that for separable (not necessarily finite-dimensional) Hilbert spaces there exists an *explicit*, universal recovery map $\mathcal{R}_{\sigma, \mathcal{N}}$ that fulfills (10) and depends only on σ and \mathcal{N} . In addition, we show that there exists a universal recovery map that satisfies (11) for finite-dimensional Hilbert spaces. We note that by the Fuchs-van de Graaf inequality [13] the fidelity can be transferred into a trace distance term such that (10) and (12) provide alternative characterizations for approximate sufficient statistics.

Result. We show that for any non-negative operator σ and for any channel \mathcal{N} there exists an *explicit* and *universal* recovery map $\mathcal{R}_{\sigma, \mathcal{N}}$ such that

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq -2 \log F(\rho, (\mathcal{R}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho)) \quad (13)$$

for all density operators ρ such that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$. A consequence of the universality of the recovery map $\mathcal{R}_{\sigma, \mathcal{N}}$ is that $(\mathcal{R}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\sigma) = \sigma$. We refer to Theorem 2.1 for a more precise statement.

Rotated and twirled Petz recovery map. We next introduce two classes of recovery maps that take the form of a rotated (or twirled) Petz recovery map. These are the building blocks for an explicit recovery map that satisfies (13). In [45], for $t \in \mathbb{R}$ the following rotated version of the Petz recovery map (given in (4)) has been introduced

$$\mathcal{R}_{\sigma, \mathcal{N}}^t : X_B \mapsto \sigma^{-it} \mathcal{P}_{\sigma, \mathcal{N}}(\mathcal{N}(\sigma)^{it} X_B \mathcal{N}(\sigma)^{-it}) \sigma^{it} . \quad (14)$$

A less specific form of a rotated Petz recovery map is given by

$$\mathcal{R}_{\sigma, \mathcal{N}}^{U, V} : X_B \mapsto V \mathcal{P}_{\sigma, \mathcal{N}}(U X_B U) V , \quad (15)$$

where U and V denote unitaries that commute with $\mathcal{N}(\sigma)$ and σ , respectively. Let $\mathcal{U}_{\sigma, A}$ be the set of unitaries on A that commute with σ . Let $\mathcal{P}(A)$ denote the set of non-negative operators on a Hilbert

space A . (If A is not finite-dimensional, we take the convention that operators in $\mathcal{P}(A)$ are trace-class.) For any $\sigma \in \mathcal{P}(A)$ and $\mathcal{N} \in \text{TPCP}(A, B)$ we denote the convex hull of rotated Petz recovery maps by

$$\mathcal{R}_{\sigma, \mathcal{N}} := \text{conv} \left(\mathcal{R}_{\sigma, \mathcal{N}}^{U, V} : U \in \mathcal{U}_{\mathcal{N}(\sigma), B} \text{ and } V \in \mathcal{U}_{\sigma, A} \right). \quad (16)$$

We note that the rotated Petz recovery map $\mathcal{R}_{\sigma, \mathcal{N}}^{U, V}$ is trace non-increasing and completely positive. For any $\sigma \in \mathcal{P}(A)$, we define

$$\mathcal{S}_\sigma(A) := \{\rho \in \mathcal{S}(A) : \text{supp}(\rho) \subseteq \text{supp}(\sigma)\}. \quad (17)$$

2 Main results

Theorem 2.1. *Let A and B be separable Hilbert spaces. For any $\sigma \in \mathcal{P}(A)$, any $\rho \in \mathcal{S}_\sigma(A)$ and any $\mathcal{N} \in \text{TPCP}(A, B)$ we have*

$$D(\rho \| \sigma) - D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) \geq -2 \log F(\rho, (\mathcal{R}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho)), \quad (18)$$

where

$$\mathcal{R}_{\sigma, \mathcal{N}}(\cdot) := \int_{\mathbb{R}} dt \beta_0(t) \mathcal{R}_{\sigma, \mathcal{N}}^{\frac{t}{2}}(\cdot) \quad (19)$$

on the support of $\mathcal{N}(\sigma)$ with $\mathcal{R}_{\sigma, \mathcal{N}}^t : X_B \mapsto \sigma^{-it} \mathcal{P}_{\sigma, \mathcal{N}}(\mathcal{N}(\sigma)^{it} X_B \mathcal{N}(\sigma)^{-it}) \sigma^{it}$ and β_0 a probability density function on \mathbb{R} defined by

$$\beta_0(t) := \frac{\pi}{2} (\cosh(\pi t) + 1)^{-1}. \quad (20)$$

If A and B are separable, $\mathcal{P}_{\sigma, \mathcal{N}}$ is defined as the adjoint of the unique linear map $\mathcal{P}_{\sigma, \mathcal{N}}^\dagger$ satisfying (3) with domain $\text{supp}(\mathcal{N}(\sigma))$ and range $\text{supp}(\sigma)$. If A and B are finite-dimensional, this unique linear map $\mathcal{P}_{\sigma, \mathcal{N}}$ is given by (4).

Figure 1 depicts the probability density β_0 as a function of $t \in \mathbb{R}$. We note that the recovery map $\mathcal{R}_{\sigma, \mathcal{N}}$ that satisfies (18) can be chosen such that it projects everything outside of the support of $\mathcal{N}(\sigma)$ to zero. For the case of finite-dimensional Hilbert spaces we can tighten the inequality (18) at the cost of losing the explicit form of the recovery map.

Theorem 2.2. *Let A and B be finite-dimensional Hilbert spaces. For any $\sigma \in \mathcal{P}(A)$ and any $\mathcal{N} \in \text{TPCP}(A, B)$ there exists a recovery map $\tilde{\mathcal{R}}_{\sigma, \mathcal{N}}$ inside the set $\mathcal{R}_{\sigma, \mathcal{N}}$ defined in (16), such that for all $\rho \in \mathcal{S}_\sigma(A)$*

$$D(\rho \| \sigma) - D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) \geq D_{\mathbb{M}}(\rho, (\tilde{\mathcal{R}}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho)) \quad (21)$$

$$\geq -2 \log F(\rho, (\tilde{\mathcal{R}}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho)). \quad (22)$$

We note that the measured relative entropy is a quantity that satisfies several desirable properties (see, e.g., [42, Lemma B.3]). Therefore the bound in (21) can be particularly helpful when it is used as a proof tool (see, e.g., Corollaries 5.2 and 5.3). Furthermore, as discussed in [7] the right-hand side of (21) can be substantially larger than (22) in some cases.

Remark 2.3 (Functoriality properties). The recovery maps $\mathcal{R}_{\sigma, \mathcal{N}}$ and $\tilde{\mathcal{R}}_{\sigma, \mathcal{N}}$ stated in Theorems 2.1 and 2.2, respectively, satisfy apart from (18) several desirable “functoriality” properties. Some of them have been stated in [47, 25, 45]. Since both recovery maps satisfy these properties, we abbreviate them both as $\mathcal{R}_{\sigma, \mathcal{N}}$ in what immediately follows.

1. **Universality.** The recovery map does not depend on ρ . This follows directly from Theorem 2.1.

2. **Perfect reconstruction of σ from $\mathcal{N}(\sigma)$.** The recovery map satisfies $(\mathcal{R}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\sigma) = \sigma$. This is clear from the fact that any rotated Petz map of the form in (14) perfectly recovers σ [45], and thus so does any convex combination of these maps. Alternatively, as the recovery map predicted by Theorem 2.1 that satisfies (18) is universal, the assertion follows by choosing $\rho = \sigma/\text{tr}(\sigma)$.
3. **Normalization.** In case $\mathcal{N} = \mathcal{I}$, where \mathcal{I} denotes the identity map, we have $\mathcal{R}_{\sigma, \mathcal{N}}(\cdot) = \Pi_\sigma(\cdot)\Pi_\sigma$. Thus if σ has full support the recovery map is equal to the identity channel. This follows directly by [45, Section 4.1] and by definition of the recovery map $\mathcal{R}_{\sigma, \mathcal{N}}(\cdot)$.
4. **Stabilization.** For any $\sigma \in \mathcal{P}(A)$, any $\mathcal{N} \in \text{TPCP}(A, B)$, any reference system R , and any full-rank $\tau \in \mathcal{P}(R)$, we have $\mathcal{R}_{\sigma \otimes \tau_{R, \mathcal{N}} \otimes \mathcal{I}_R}(\cdot) = \mathcal{R}_{\sigma, \mathcal{N}} \otimes \mathcal{I}_R(\cdot)$. This follows by combining [45, Section 4.2] together with the normalization property discussed above.

We note that by following [45, Sections 4.2 and 4.3] together with the strengthening of Theorem 2.1 given in (23) we can obtain remainder terms that fulfill some parallel and serial composition rules.

The proof of Theorem 2.1 consists of two parts. We first prove the statement for finite-dimensional Hilbert spaces A and B by employing a strengthened version of Hadamard's three-line theorem that is due to Hirschman [18]. By an approximation argument we show that the result remains valid for separable Hilbert spaces. Theorem 2.2 is proven differently. Via Sion's minimax theorem [41, Corollary 3.3], we show the existence of a universal recovery map that satisfies (21) for any finite set of states $\rho \in \mathcal{S}(A)$ (see Step 1). We then show, by using the concept of an ε -net, how this implies the existence of a universal recovery map that satisfies (21) for all $\rho \in \mathcal{S}(A)$.

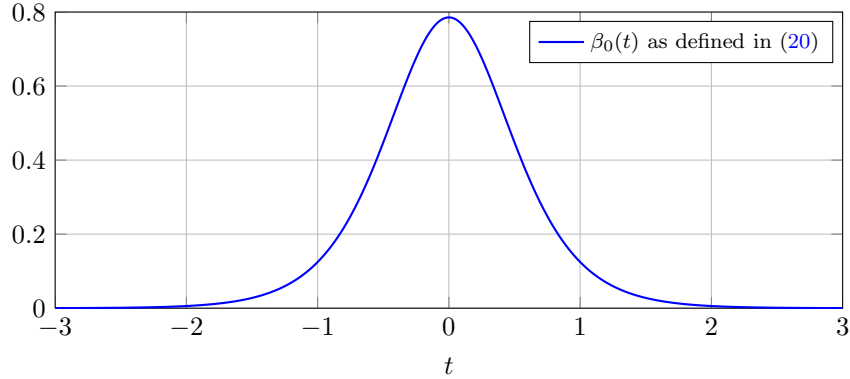


Figure 1: This plot depicts the probability density β_0 defined in (20) as a function of $t \in \mathbb{R}$. We see that it is peaked around $t = 0$ which corresponds to the Petz recovery map, i.e., $\mathcal{R}_{\sigma, \mathcal{N}}^{t=0} = \mathcal{P}_{\sigma, \mathcal{N}}$.

3 Proof of Theorem 2.1

Step 1: Proof for finite-dimensional Hilbert spaces

In this step we assume that the Hilbert spaces A and B are finite-dimensional. We note that we will prove a slightly stronger version of Theorem 2.1 that by concavity of the logarithm and the fidelity immediately implies Theorem 2.1.

Stronger version of Theorem 2.1. *For any $\sigma \in \mathcal{P}(A)$, any $\rho \in \mathcal{S}_\sigma(A)$ and any $\mathcal{N} \in \text{TPCP}(A, B)$ we have*

$$D(\rho \| \sigma) - D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) \geq -2 \int_{\mathbb{R}} dt \beta_0(t) \log F(\rho, (\mathcal{R}_{\sigma, \mathcal{N}}^{\frac{t}{2}} \circ \mathcal{N})(\rho)) , \quad (23)$$

where $\mathcal{R}_{\sigma, \mathcal{N}}^t(\cdot)$ is defined in (14) and the probability density function $\beta_0(t) := \frac{\pi}{2}(\cosh(\pi t) + 1)^{-1}$.

Remark 3.1. Inequality (23) together with the fact that the mapping $t \mapsto \mathcal{R}_{\sigma, \mathcal{N}}^t$ is continuous implies that for any $\sigma \in \mathcal{P}(A)$, $\rho \in \mathcal{S}_\sigma(A)$ and $\mathcal{N} \in \text{TPCP}(A, B)$ such that $D(\rho \parallel \sigma) = D(\mathcal{N}(\rho) \parallel \mathcal{N}(\sigma))$ we have $(\mathcal{R}_{\sigma, \mathcal{N}}^t \circ \mathcal{N})(\rho) = \rho$ and $(\mathcal{R}_{\sigma, \mathcal{N}}^t \circ \mathcal{N})(\sigma) = \sigma$ for all $t \in \mathbb{R}$ with $\mathcal{R}_{\sigma, \mathcal{N}}^t(\cdot)$ defined in (14). This follows because $F(\omega, \tau) \in [0, 1]$ and $F(\omega, \tau) = 1$ if and only if $\omega = \tau$ for density operators ω and τ .

Our proof of (23) is similar to the approach taken in [45]. There are two main ingredients: a Rényi generalization of a relative entropy difference [39] and Hirschman's improvement of the Hadamard three-line theorem [18]. We begin by recalling these two ingredients and then proceed to a proof of (23).

Let $L(A)$ denote the space of bounded linear operators acting on a Hilbert space A . For any $L \in L(A)$ the *Schatten p -norm* is defined as

$$\|L\|_p := (\text{tr}(|L|^p))^{\frac{1}{p}} \quad \text{for } p \in [1, \infty), \quad (24)$$

where $|L| := \sqrt{L^\dagger L}$. A Rényi generalization of a relative entropy difference is defined as [39]

$$\tilde{\Delta}_\alpha(\rho, \sigma, \mathcal{N}) := \frac{2\alpha}{\alpha - 1} \log \left\| \left([\mathcal{N}(\rho)]^{\frac{1-\alpha}{2\alpha}} [\mathcal{N}(\sigma)]^{\frac{\alpha-1}{2\alpha}} \otimes \text{id}_E \right) U_{A \rightarrow BE} \sigma^{\frac{1-\alpha}{2\alpha}} \rho^{\frac{1}{2}} \right\|_{2\alpha}, \quad (25)$$

where $\alpha \in (0, 1) \cup (1, \infty)$, and $U_{A \rightarrow BE}$ is an isometric extension of the channel \mathcal{N} . That is, $U_{A \rightarrow BE}$ is a linear isometry satisfying $\text{tr}_E(U_{A \rightarrow BE}(\cdot)U_{A \rightarrow BE}^\dagger) = \mathcal{N}(\cdot)$ and $U_{A \rightarrow BE}^\dagger U_{A \rightarrow BE} = \text{id}_A$. All isometric extensions of a channel are related by an isometry acting on the environment system E , so that the definition in (25) is invariant under any such choice. Recall also that the adjoint \mathcal{N}^\dagger of a channel is given in terms of an isometric extension U as $\mathcal{N}^\dagger(\cdot) = U^\dagger(\cdot \otimes \text{id}_E)U$.

Lemma 3.2 ([39, 45]). *The following limit holds for ρ , σ , and \mathcal{N} as given in the statement of Theorem 2.1:*

$$\lim_{\alpha \rightarrow 1} \tilde{\Delta}_\alpha(\rho, \sigma, \mathcal{N}) = D(\rho \parallel \sigma) - D(\mathcal{N}(\rho) \parallel \mathcal{N}(\sigma)). \quad (26)$$

For $\alpha = \frac{1}{2}$, observe that

$$\tilde{\Delta}_{\frac{1}{2}}(\rho, \sigma, \mathcal{N}) = -2 \log \left\| \left([\mathcal{N}(\rho)]^{\frac{1}{2}} [\mathcal{N}(\sigma)]^{-\frac{1}{2}} \otimes \text{id}_E \right) U_{A \rightarrow BE} \sigma^{\frac{1}{2}} \rho^{\frac{1}{2}} \right\|_1 = -2 \log F\left(\rho, \mathcal{P}_{\sigma, \mathcal{N}}(\mathcal{N}(\rho))\right). \quad (27)$$

The following lemma is based on Hirschman's improvement of the Hadamard three-line theorem [18], and for completeness, we provide a proof in Appendix A.

Lemma 3.3. *Let $S := \{z \in \mathbb{C} : 0 \leq \text{Re}\{z\} \leq 1\}$ and let $G : S \rightarrow L(\mathcal{H})$ be a bounded map that is holomorphic on the interior of S and continuous on the boundary. Let $\theta \in (0, 1)$ and define p_θ by*

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \quad (28)$$

where $p_0, p_1 \in [1, \infty]$. Then the following bound holds

$$\log(\|G(\theta)\|_{p_\theta}) \leq \int_{\mathbb{R}} dt \left(\alpha_\theta(t) \log(\|G(it)\|_{p_0}^{1-\theta}) + \beta_\theta(t) \log(\|G(1+it)\|_{p_1}^\theta) \right), \quad (29)$$

where $\alpha_\theta(t)$ and $\beta_\theta(t)$ are defined by

$$\alpha_\theta(t) := \frac{\sin(\pi\theta)}{2(1-\theta)(\cosh(\pi t) - \cos(\pi\theta))} \quad \text{and} \quad \beta_\theta(t) := \frac{\sin(\pi\theta)}{2\theta(\cosh(\pi t) + \cos(\pi\theta))}. \quad (30)$$

Remark 3.4. Fix $\theta \in (0, 1)$. Observe that $\alpha_\theta(t), \beta_\theta(t) \geq 0$ for all $t \in \mathbb{R}$ and we have

$$\int_{\mathbb{R}} dt \alpha_\theta(t) = \int_{\mathbb{R}} dt \beta_\theta(t) = 1, \quad (31)$$

(see, e.g., [16, Exercise 1.3.8]) so that $\alpha_\theta(t)$ and $\beta_\theta(t)$ can be interpreted as probability density functions. Furthermore, the following limit holds

$$\lim_{\theta \searrow 0} \beta_\theta(t) = \frac{\pi}{2(\cosh(\pi t) + 1)} = \beta_0(t) , \quad (32)$$

where β_0 is also a probability density function on \mathbb{R} .

We can now readily establish the desired result in (23). In what follows, we abbreviate the isometric extension $U_{A \rightarrow BE}$ of the channel \mathcal{N} as U . Pick

$$G(z) := \left([\mathcal{N}(\rho)]^{\frac{\theta}{2}} [\mathcal{N}(\sigma)]^{-\frac{\theta}{2}} \otimes \text{id}_E \right) U \sigma^{\frac{\theta}{2}} \rho^{\frac{1}{2}} , \quad (33)$$

$p_0 = 2$, $p_1 = 1$, and $\theta \in (0, 1)$, which fixes $p_\theta = \frac{2}{1+\theta}$. The operator valued-function $G(z)$ satisfies the conditions needed to apply Lemma 3.3. For the choices above, we find that

$$\|G(\theta)\|_{\frac{2}{1+\theta}} = \left\| \left([\mathcal{N}(\rho)]^{\frac{\theta}{2}} [\mathcal{N}(\sigma)]^{-\frac{\theta}{2}} \otimes \text{id}_E \right) U \sigma^{\frac{\theta}{2}} \rho^{\frac{1}{2}} \right\|_{\frac{2}{1+\theta}} , \quad (34)$$

and

$$\|G(it)\|_2 = \left\| \left([\mathcal{N}(\rho)]^{\frac{it}{2}} [\mathcal{N}(\sigma)]^{-\frac{it}{2}} \otimes \text{id}_E \right) U \sigma^{it} \rho^{\frac{1}{2}} \right\|_2 \leq \left\| \rho^{\frac{1}{2}} \right\|_2 = 1 , \quad (35)$$

as well as

$$\begin{aligned} \|G(1+it)\|_1 &= \left\| \left([\mathcal{N}(\rho)]^{\frac{1+it}{2}} [\mathcal{N}(\sigma)]^{-\frac{1+it}{2}} \otimes \text{id}_E \right) U \sigma^{\frac{1+it}{2}} \rho^{\frac{1}{2}} \right\|_1 \\ &= \left\| \left([\mathcal{N}(\rho)]^{\frac{it}{2}} [\mathcal{N}(\rho)]^{\frac{1}{2}} [\mathcal{N}(\sigma)]^{-\frac{it}{2}} [\mathcal{N}(\sigma)]^{-\frac{1}{2}} \otimes \text{id}_E \right) U \sigma^{\frac{1}{2}} \sigma^{\frac{it}{2}} \rho^{\frac{1}{2}} \right\|_1 \\ &= \left\| \left([\mathcal{N}(\rho)]^{\frac{1}{2}} [\mathcal{N}(\sigma)]^{-\frac{it}{2}} [\mathcal{N}(\sigma)]^{-\frac{1}{2}} \otimes \text{id}_E \right) U \sigma^{\frac{1}{2}} \sigma^{\frac{it}{2}} \rho^{\frac{1}{2}} \right\|_1 \\ &= F(\rho, (\mathcal{R}_{\sigma, \mathcal{N}}^{\frac{it}{2}} \circ \mathcal{N})(\rho)) . \end{aligned} \quad (36)$$

Then we can apply the fact that $\|G(it)\|_2 \leq 1$ and (29) to conclude that the following bound holds for all $\theta \in (0, 1)$

$$\log \left\| \left([\mathcal{N}(\rho)]^{\frac{\theta}{2}} [\mathcal{N}(\sigma)]^{-\frac{\theta}{2}} \otimes \text{id}_E \right) U \sigma^{\frac{\theta}{2}} \rho^{\frac{1}{2}} \right\|_{\frac{2}{1+\theta}} \leq \int_{\mathbb{R}} dt \beta_\theta(t) \log \left(F(\rho, (\mathcal{R}_{\sigma, \mathcal{N}}^{\frac{t}{2}} \circ \mathcal{N})(\rho))^\theta \right) , \quad (37)$$

which implies

$$-\frac{2}{\theta} \log \left\| \left([\mathcal{N}(\rho)]^{\frac{\theta}{2}} [\mathcal{N}(\sigma)]^{-\frac{\theta}{2}} \otimes \text{id}_E \right) U \sigma^{\frac{\theta}{2}} \rho^{\frac{1}{2}} \right\|_{\frac{2}{1+\theta}} \geq -2 \int_{\mathbb{R}} dt \beta_\theta(t) \log F(\rho, (\mathcal{R}_{\sigma, \mathcal{N}}^{\frac{t}{2}} \circ \mathcal{N})(\rho)) . \quad (38)$$

Letting $\theta = \frac{1-\alpha}{\alpha}$, we see that this is the same as

$$\tilde{\Delta}_\alpha(\rho, \sigma, \mathcal{N}) \geq -2 \int_{\mathbb{R}} dt \beta_{\frac{1-\alpha}{\alpha}}(t) \log F(\rho, (\mathcal{R}_{\sigma, \mathcal{N}}^{\frac{t}{2}} \circ \mathcal{N})(\rho)) . \quad (39)$$

Since the inequality in (38) holds for all $\theta \in (0, 1)$ and thus (39) holds for all $\alpha \in (\frac{1}{2}, 1)$, we can take the limit as $\alpha \nearrow 1$ and apply (26), (32), and the dominated convergence theorem to conclude that (23) holds.

Remark 3.5. If A and B are finite-dimensional Hilbert spaces the statement in Theorem 2.1 can be slightly generalized. For any $\sigma \geq 0$ and any trace non-increasing completely positive map \mathcal{N} with Kraus operators $\{N_i\}$ such that $0 \neq \sum_i N_i^\dagger N_i \leq \text{id}$ the recovery map $\mathcal{R}_{\sigma, \mathcal{N}}$ defined in Theorem 2.1 satisfies (18) for all subnormalized density operators $\rho \geq 0$ with $0 < \text{tr}(\rho) \leq 1$ such that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$. This follows by the same argument given in Step 1 by using $U_{A \rightarrow BE} = \sum_i N_i \otimes |i\rangle_E$ as an isometric extension of the trace non-increasing and completely positive map \mathcal{N} .

Step 2: Extension to infinite dimensions

In this step the Hilbert spaces A and B are assumed to be separable (not necessarily finite-dimensional). We will show how to lift the stronger version of Theorem 2.1 for finite-dimensional Hilbert spaces, such that it can apply to states ρ and σ and a channel \mathcal{N} associated with separable Hilbert spaces. This is accomplished via a limiting argument. Let $\{\Pi_A^a\}_{a \in \mathbb{N}}$ and $\{\Pi_B^b\}_{b \in \mathbb{N}}$ be sequences of finite-rank projectors on A and B , respectively, that converge to id_A and id_B , respectively, with respect to the weak operator topology, meaning that

$$\lim_{a \rightarrow \infty} \langle \psi | \Pi_A^a | \phi \rangle = \langle \psi | \phi \rangle \quad (40)$$

for all vectors $|\phi\rangle, |\psi\rangle \in A$ (similarly for $\Pi_B^b \rightarrow \text{id}_B$). For $\sigma \in \mathcal{P}(A)$ and $\rho \in \mathcal{S}_\sigma(A)$ we consider projected versions

$$\sigma^a := \Pi_A^a \sigma \Pi_A^a \quad \text{and} \quad \rho^a := \Pi_A^a \rho \Pi_A^a . \quad (41)$$

We note that the sequences $\{\rho^a\}_{a \in \mathbb{N}}$ and $\{\sigma^a\}_{a \in \mathbb{N}}$ converge to ρ and σ , respectively, in the trace norm (see, e.g., Corollary 2 of [15] or Lemma 11.1 of [19]). Let \mathcal{S}^a be the set of non-negative operators that is generated by (41) for all $\rho \in \mathcal{S}$. For any $\mathcal{N} \in \text{TPCP}(A, B)$ we define its analogue with a projection at the input and output as

$$\mathcal{N}^{a,b}(\cdot) := \Pi_B^b \mathcal{N}(\Pi_A^a(\cdot) \Pi_A^a) \Pi_B^b . \quad (42)$$

Note that $\mathcal{N}^{a,b}$ converges to \mathcal{N} in the weak operator topology, in the sense that

$$\lim_{a,b \rightarrow \infty} \langle \phi' | \mathcal{N}^{a,b}(|\phi\rangle\langle\psi|) | \psi' \rangle = \langle \phi' | \mathcal{N}(|\phi\rangle\langle\psi|) | \psi' \rangle, \quad (43)$$

for all vectors $|\phi\rangle, |\psi\rangle \in A$ and $|\phi'\rangle, |\psi'\rangle \in B$ (cf., the discussions in [40]).

We start by proving two lemmas that show how the difference of relative entropies and the fidelity, respectively, change when considering projected states.

Lemma 3.6. *For any $\sigma \in \mathcal{P}(A)$, any $\rho \in \mathcal{S}_\sigma(A)$, and any $\mathcal{N} \in \text{TPCP}(A, B)$, we have*

$$\limsup_{a \rightarrow \infty} \limsup_{b \rightarrow \infty} D(\rho^a \| \sigma^a) - D(\mathcal{N}^{a,b}(\rho^a) \| \mathcal{N}^{a,b}(\sigma^a)) \leq D(\rho \| \sigma) - D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) . \quad (44)$$

Proof. By Lemma 3 in [28] and since the relative entropy is lower semicontinuous [19, Exercise 7.22], we obtain

$$D(\rho^a \| \sigma^a) \leq D(\rho \| \sigma) \leq \liminf_{a \rightarrow \infty} D(\rho^a \| \sigma^a) , \quad (45)$$

which implies that

$$\lim_{a \rightarrow \infty} D(\rho^a \| \sigma^a) = D(\rho \| \sigma) . \quad (46)$$

The lower semicontinuity of the relative entropy implies that

$$\liminf_{a \rightarrow \infty} \liminf_{b \rightarrow \infty} D(\mathcal{N}^{a,b}(\rho^a) \| \mathcal{N}^{a,b}(\sigma^a)) \geq \liminf_{a \rightarrow \infty} D(\mathcal{N}(\rho^a) \| \mathcal{N}(\sigma^a)) \quad (47)$$

$$\geq D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) . \quad (48)$$

This proves the assertion. \square

Lemma 3.7. *The Petz recovery map $\mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}$ defined in (4) satisfies $\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}} = \mathcal{P}_{\sigma, \mathcal{N}}$ (in the sense of (43)), where $\mathcal{P}_{\sigma, \mathcal{N}}$ is a recovery map that is defined as the (unique) linear map with domain $\text{supp}(\mathcal{N}(\sigma))$ and range $\text{supp}(\sigma)$ satisfying*

$$\langle a_2, \mathcal{N}^\dagger(a_1) \rangle_\sigma = \langle \mathcal{P}_{\sigma, \mathcal{N}}^\dagger(a_2), a_1 \rangle_{\mathcal{N}(\sigma)} \quad \text{for all } a_1 \in \mathcal{L}(B), a_2 \in \mathcal{L}(A) , \quad (49)$$

with the weighted inner product $\langle a, b \rangle_\omega := \text{tr}(a^\dagger \omega^{\frac{1}{2}} b \omega^{\frac{1}{2}})$ and ω positive semi-definite and trace class.

Proof. We begin by outlining the proof. As shown by Petz [34, 35, 36] (see also [33, Chapter 8]), the map $\mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}$ defined in (4) is the unique linear map with domain $\text{supp}(\mathcal{N}^{a,b}(\sigma^a))$ and range $\text{supp}(\sigma^a)$ satisfying

$$\langle a_2, (\mathcal{N}^{a,b})^\dagger(a_1) \rangle_{\sigma^a} = \left\langle \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2), a_1 \right\rangle_{\mathcal{N}^{a,b}(\sigma^a)} \quad \text{for all } a_1 \in \text{L}(B), a_2 \in \text{L}(A), \quad (50)$$

and the map $\mathcal{P}_{\sigma, \mathcal{N}}$ is the unique linear map with domain $\text{supp}(\mathcal{N}(\sigma))$ and range $\text{supp}(\sigma)$ satisfying

$$\langle a_2, (\mathcal{N})^\dagger(a_1) \rangle_\sigma = \left\langle \mathcal{P}_{\sigma, \mathcal{N}}^\dagger(a_2), a_1 \right\rangle_{\mathcal{N}(\sigma)} \quad \text{for all } a_1 \in \text{L}(B), a_2 \in \text{L}(A). \quad (51)$$

We will first show that

$$\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \langle a_2, (\mathcal{N}^{a,b})^\dagger(a_1) \rangle_{\sigma^a} = \langle a_2, \mathcal{N}^\dagger(a_1) \rangle_\sigma, \quad (52)$$

for all $a_1 \in \text{L}(B), a_2 \in \text{L}(A)$, which by (50) and (51) implies that

$$\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \left\langle \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2), a_1 \right\rangle_{\mathcal{N}^{a,b}(\sigma^a)} = \left\langle \mathcal{P}_{\sigma, \mathcal{N}}^\dagger(a_2), a_1 \right\rangle_{\mathcal{N}(\sigma)} \quad (53)$$

for all $a_1 \in \text{L}(B), a_2 \in \text{L}(A)$. After showing that

$$\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \left\langle \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2), a_1 \right\rangle_{\mathcal{N}^{a,b}(\sigma^a)} = \lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \left\langle \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2), a_1 \right\rangle_{\mathcal{N}(\sigma)}, \quad (54)$$

we can conclude from (53) and (54) that

$$\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \left\langle \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2), a_1 \right\rangle_{\mathcal{N}(\sigma)} = \left\langle \mathcal{P}_{\sigma, \mathcal{N}}^\dagger(a_2), a_1 \right\rangle_{\mathcal{N}(\sigma)}. \quad (55)$$

From there, we argue that this implies the convergence $\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}} = \mathcal{P}_{\sigma, \mathcal{N}}$, in the sense of (43).

Thus, we need to establish (52) and (54), and we begin by proving (52). Consider that

$$\langle a_2, (\mathcal{N}^{a,b})^\dagger(a_1) \rangle_\sigma = \text{tr} \left(a_2^\dagger (\sigma^a)^{\frac{1}{2}} (\mathcal{N}^{a,b})^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} \right). \quad (56)$$

Then to prove (52), we need to show that the following limit holds for all bounded a_1 and a_2 :

$$\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \left| \text{tr} \left(a_2^\dagger (\sigma^a)^{\frac{1}{2}} (\mathcal{N}^{a,b})^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} \right) - \text{tr} \left(a_2^\dagger \sigma^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} \right) \right| = 0. \quad (57)$$

From the fact that any bounded square operator can be written as a sum of a real part and an imaginary part, each which of in turn can be written as a sum of a positive and negative part [23, Chapter 2], it suffices to prove the limit in (57) for positive semi-definite a_1 and a_2 . So we take a_1 and a_2 positive semi-definite in what follows. Consider that

$$\begin{aligned} & \left| \text{tr} \left(a_2 (\sigma^a)^{\frac{1}{2}} (\mathcal{N}^{a,b})^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} \right) - \text{tr} \left(a_2 \sigma^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} \right) \right| \\ & \leq \left| \text{tr} \left(a_2 \left[(\sigma^a)^{\frac{1}{2}} (\mathcal{N}^{a,b})^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} - (\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} \right] \right) \right| \\ & \quad + \left| \text{tr} \left(a_2 \left[(\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} - \sigma^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} \right] \right) \right|. \end{aligned} \quad (58)$$

We begin with the first term on the right of (58):

$$\begin{aligned} & \left| \text{tr} \left(a_2 \left[(\sigma^a)^{\frac{1}{2}} (\mathcal{N}^{a,b})^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} - (\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} \right] \right) \right| \\ & = \left| \text{tr} \left(\left(\mathcal{N}^{a,b} \left[(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right] - \mathcal{N} \left[(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right] \right) a_1 \right) \right| \end{aligned} \quad (59)$$

$$\leq \left\| \mathcal{N}^{a,b} \left[(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right] - \mathcal{N} \left[(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right] \right\|_1 \|a_1\|_\infty, \quad (60)$$

where we applied the Hölder inequality in the last line. So now we focus on bounding the trace norm above. Consider that $(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}}$ is a trace-class operator because

$$\left\| (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right\|_1 \leq \left\| (\sigma^a)^{\frac{1}{2}} \right\|_2 \|a_2\|_\infty \left\| (\sigma^a)^{\frac{1}{2}} \right\|_2 \quad (61)$$

$$= \|\sigma^a\|_1 \|a_2\|_\infty \quad (62)$$

$$\leq \|\sigma\|_1 \|a_2\|_\infty. \quad (63)$$

So then

$$\begin{aligned} & \left\| \mathcal{N}^{a,b} \left[(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right] - \mathcal{N} \left[(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right] \right\|_1 \\ &= \left\| \Pi_B^b \mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] \Pi_B^b - \mathcal{N} \left[(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right] \right\|_1 \end{aligned} \quad (64)$$

$$\begin{aligned} & \leq \left\| \Pi_B^b \mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] \Pi_B^b - \mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] \right\|_1 \\ & \quad + \left\| \mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] - \mathcal{N} \left[(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right] \right\|_1. \end{aligned} \quad (65)$$

We now bound the last two terms. Let

$$\omega_B^a := \frac{\mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right]}{\text{tr} \left(\mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] \right)}. \quad (66)$$

Then the first term is bounded as

$$\begin{aligned} & \left\| \Pi_B^b \mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] \Pi_B^b - \mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] \right\|_1 \\ &= \text{tr} \left(\mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] \right) \left\| \Pi_B^b \omega^a \Pi_B^b - \omega^a \right\|_1 \end{aligned} \quad (67)$$

$$\leq \text{tr} \left(\mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] \right) \left[2\sqrt{1 - \text{tr}(\Pi_B^b \omega^a)} \right], \quad (68)$$

where the last line follows from a standard result (see, e.g., [42, Lemma A.2]). Taking the limit as $b \rightarrow \infty$ and as $a \rightarrow \infty$ then makes this term arbitrarily small. For the other term, consider that

$$\begin{aligned} & \left\| \mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] - \mathcal{N} \left[(\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right] \right\|_1 \leq \left\| \Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a - (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \right\|_1 \\ &= 0, \end{aligned} \quad (69)$$

where the last equality follows because $\Pi_A^a (\sigma^a)^{\frac{1}{2}} = (\sigma^a)^{\frac{1}{2}}$.

Now we handle the second term on the right in (58):

$$\begin{aligned} & \left| \text{tr} \left(a_2 \left[(\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} - \sigma^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} \right] \right) \right| \\ & \leq \left| \text{tr} \left(a_2 \left[(\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} - (\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} \right] \right) \right| \\ & \quad + \left| \text{tr} \left(a_2 \left[(\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} - \sigma^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} \right] \right) \right|. \end{aligned} \quad (70)$$

The first term on the right in (70) is bounded as follows:

$$\begin{aligned} & \left| \text{tr} \left(a_2 \left[(\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) (\sigma^a)^{\frac{1}{2}} - (\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} \right] \right) \right| \\ &= \left| \text{tr} \left(a_2 (\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \left[(\sigma^a)^{\frac{1}{2}} - \sigma^{\frac{1}{2}} \right] \right) \right| \end{aligned} \quad (71)$$

$$\leq \left[\text{tr} (a_2^2 \sigma^a) \text{tr} \left([\mathcal{N}^\dagger(a_1)]^2 \left[(\sigma^a)^{\frac{1}{2}} - \sigma^{\frac{1}{2}} \right]^2 \right) \right]^{\frac{1}{2}} \quad (72)$$

$$\leq \left[\|a_2\|_\infty^2 \|a_1\|_\infty^2 \text{tr}(\sigma^a) \text{tr} \left(\left[(\sigma^a)^{\frac{1}{2}} - \sigma^{\frac{1}{2}} \right]^2 \right) \right]^{\frac{1}{2}} \quad (73)$$

$$\leq \|a_2\|_\infty \|a_1\|_\infty \|\sigma^a\|_1^{\frac{1}{2}} \left\| (\sigma^a)^{\frac{1}{2}} - \sigma^{\frac{1}{2}} \right\|_2 \quad (74)$$

$$\leq \|a_2\|_\infty \|a_1\|_\infty \|\sigma\|_1^{\frac{1}{2}} [\|\sigma^a - \sigma\|_1]^{\frac{1}{2}}. \quad (75)$$

The first inequality follows from the Cauchy-Schwarz inequality. The last inequality is a consequence of the Powers-Stormer inequality. By essentially the same reasoning, we obtain the following bound on the other term on the right in (70):

$$\left| \text{tr} \left(a_2 \left[(\sigma^a)^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} - \sigma^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}} \right] \right) \right| \leq \|a_2\|_\infty \|a_1\|_\infty \|\sigma\|_1^{\frac{1}{2}} [\|\sigma^a - \sigma\|_1]^{\frac{1}{2}}. \quad (76)$$

Putting everything together, we find that

$$\begin{aligned} & \left| \text{tr} (a_2^\dagger (\sigma^a)^{\frac{1}{2}} (\mathcal{N}^{a,b})^\dagger(a_1) (\sigma^a)^{\frac{1}{2}}) - \text{tr} (a_2^\dagger \sigma^{\frac{1}{2}} \mathcal{N}^\dagger(a_1) \sigma^{\frac{1}{2}}) \right| \\ & \leq \text{tr} \left(\mathcal{N} \left[\Pi_A^a (\sigma^a)^{\frac{1}{2}} a_2 (\sigma^a)^{\frac{1}{2}} \Pi_A^a \right] \right) \left[2\sqrt{1 - \text{tr}(\Pi_B^b \omega^a)} \right] + 2 \|a_2\|_\infty \|a_1\|_\infty \|\sigma\|_1^{1/2} [\|\sigma^a - \sigma\|_1]^{\frac{1}{2}}, \end{aligned} \quad (77)$$

from which we can conclude (52).

We now prove (54), by a reasoning that is very similar to the above. Consider that

$$\begin{aligned} & \left| \langle \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2), a_1 \rangle_{\mathcal{N}^{a,b}(\sigma^a)} - \langle \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2), a_1 \rangle_{\mathcal{N}(\sigma)} \right| \\ &= \left| \text{tr} \left(\mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2) \left[\mathcal{N}^{a,b}(\sigma^a)^{\frac{1}{2}} a_1 \mathcal{N}^{a,b}(\sigma^a)^{\frac{1}{2}} - \mathcal{N}(\sigma)^{\frac{1}{2}} a_1 \mathcal{N}(\sigma)^{\frac{1}{2}} \right] \right) \right| \end{aligned} \quad (78)$$

$$\begin{aligned} & \leq \left| \text{tr} \left(\mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2) \mathcal{N}^{a,b}(\sigma^a)^{\frac{1}{2}} a_1 \left[\mathcal{N}^{a,b}(\sigma^a)^{\frac{1}{2}} - \mathcal{N}(\sigma)^{\frac{1}{2}} \right] \right) \right| \\ & \quad + \left| \text{tr} \left(\mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2) \left[\mathcal{N}^{a,b}(\sigma^a)^{\frac{1}{2}} - \mathcal{N}(\sigma)^{\frac{1}{2}} \right] a_1 \mathcal{N}(\sigma)^{\frac{1}{2}} \right) \right|. \end{aligned} \quad (79)$$

By the same methods as given above, we find that

$$\begin{aligned} & \left| \text{tr} \left(\mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2) \mathcal{N}^{a,b}(\sigma^a)^{\frac{1}{2}} a_1 \left[\mathcal{N}^{a,b}(\sigma^a)^{\frac{1}{2}} - \mathcal{N}(\sigma)^{\frac{1}{2}} \right] \right) \right| \\ & \leq \|a_2\|_\infty \|a_1\|_\infty \|\sigma\|_1^{\frac{1}{2}} [\|\mathcal{N}^{a,b}(\sigma^a) - \mathcal{N}(\sigma)\|_1]^{\frac{1}{2}}, \end{aligned} \quad (80)$$

and

$$\begin{aligned} & \left| \text{tr} \left(\mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2) \left[\mathcal{N}^{a,b}(\sigma^a)^{\frac{1}{2}} - \mathcal{N}(\sigma)^{\frac{1}{2}} \right] a_1 \mathcal{N}(\sigma)^{\frac{1}{2}} \right) \right| \\ & \leq \|a_2\|_\infty \|a_1\|_\infty \|\sigma\|_1^{\frac{1}{2}} [\|\mathcal{N}^{a,b}(\sigma^a) - \mathcal{N}(\sigma)\|_1]^{\frac{1}{2}}. \end{aligned} \quad (81)$$

Finally, consider that

$$\|\mathcal{N}^{a,b}(\sigma^a) - \mathcal{N}(\sigma)\|_1 \leq \|\mathcal{N}^{a,b}(\sigma^a) - \mathcal{N}^{a,b}(\sigma)\|_1 + \|\mathcal{N}^{a,b}(\sigma) - \mathcal{N}(\sigma)\|_1 \quad (82)$$

$$\leq \|\sigma^a - \sigma\|_1 + \|\Pi^b \mathcal{N}(\Pi^a \sigma \Pi^a) \Pi^b - \mathcal{N}(\Pi^a \sigma \Pi^a)\|_1 + \|\mathcal{N}(\Pi^a \sigma \Pi^a) - \mathcal{N}(\sigma)\|_1 \quad (83)$$

$$\leq \|\sigma^a - \sigma\|_1 + \|\Pi^b \mathcal{N}(\Pi^a \sigma \Pi^a) \Pi^b - \mathcal{N}(\Pi^a \sigma \Pi^a)\|_1 + \|\Pi^a \sigma \Pi^a - \sigma\|_1, \quad (84)$$

which converges also in the limit as $b \rightarrow \infty$ and as $a \rightarrow \infty$. Putting everything together, we find that

$$\begin{aligned} & \left| \langle \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2), a_1 \rangle_{\mathcal{N}^{a,b}(\sigma^a)} - \langle \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(a_2), a_1 \rangle_{\mathcal{N}(\sigma)} \right| \\ & \leq 2 \|a_2\|_\infty \|a_1\|_\infty \|\sigma\|_1^{\frac{1}{2}} \left[2 \|\sigma^a - \sigma\|_1 + \|\Pi^b \mathcal{N}(\Pi^a \sigma \Pi^a) \Pi^b - \mathcal{N}(\Pi^a \sigma \Pi^a)\|_1 \right]^{\frac{1}{2}}, \end{aligned} \quad (85)$$

which allows us to conclude (54), and in turn, (55).

It remains to argue for the convergence $\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}} = \mathcal{P}_{\sigma, \mathcal{N}}$, in the sense of (43). Consider that (55) implies that

$$\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \langle \phi' | \mathcal{N}(\sigma)^{\frac{1}{2}} \mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger(|\phi\rangle\langle\psi|) \mathcal{N}(\sigma)^{\frac{1}{2}} | \psi' \rangle = \langle \phi' | \mathcal{N}(\sigma)^{\frac{1}{2}} \mathcal{P}_{\sigma, \mathcal{N}}^\dagger(|\phi\rangle\langle\psi|) \mathcal{N}(\sigma)^{\frac{1}{2}} | \psi' \rangle, \quad (86)$$

for all $|\phi\rangle, |\psi\rangle \in A$ and $|\phi'\rangle, |\psi'\rangle \in B$. This establishes convergence of $\mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}^\dagger$ to $\mathcal{P}_{\sigma, \mathcal{N}}^\dagger$ for all $|\phi\rangle, |\psi\rangle \in A$ and $|\phi'\rangle, |\psi'\rangle \in \text{supp}(\mathcal{N}(\sigma))$, which in turn allows us to conclude convergence of $\mathcal{P}_{\sigma^a, \mathcal{N}^{a,b}}$ to $\mathcal{P}_{\sigma, \mathcal{N}}$. \square

Before stating the following lemma, we introduce the shorthand

$$\mathcal{U}_{\omega, t} : X \mapsto \omega^{it} X \omega^{-it}, \quad (87)$$

where ω is a positive semi-definite operator.

Lemma 3.8. *For any $\sigma \in \mathcal{P}(A)$, any $\rho \in \mathcal{S}_\sigma(A)$, any $\mathcal{N} \in \text{TPCP}(A, B)$, and all $t \in \mathbb{R}$, the following limit holds*

$$\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} F(\rho^a, (\mathcal{R}_{\sigma^a, \mathcal{N}^{a,b}}^t \circ \mathcal{N}^{a,b})(\rho^a)) = F(\rho, (\mathcal{R}_{\sigma, \mathcal{N}}^t \circ \mathcal{N})(\rho)). \quad (88)$$

Proof. Consider that we have the following convergences:

$$\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \mathcal{U}_{\mathcal{N}^{a,b}(\sigma^a), t} = \mathcal{U}_{\mathcal{N}(\sigma), t} \quad \text{and} \quad \lim_{a \rightarrow \infty} \mathcal{U}_{\sigma^a, t} = \mathcal{U}_{\sigma, t}, \quad (89)$$

each understood to be in the weak sense in (43). The serial concatenation of weakly converging channels converges to the serial concatenation of the limit, so that $\lim_{a \rightarrow \infty} \lim_{b \rightarrow \infty} \mathcal{R}_{\sigma^a, \mathcal{N}^{a,b}}^t \circ \mathcal{N}^{a,b} = \mathcal{R}_{\sigma, \mathcal{N}}^t \circ \mathcal{N}$, where we used the above and Lemma 3.7. Then we can conclude that the fidelity converges because it is continuous in its inputs (see, e.g., [11, Lemma B.9]), and for our case considered here, convergence in the weak sense implies convergence in the trace norm [19, Chapter 11]. This proves the assertion. \square

By invoking the stronger version of Theorem 2.1 for finite-dimensional Hilbert spaces (that has been proven in Step 1)¹ together with Lemmas 3.6 and 3.8, the fact that any separable Hilbert space is isomorphic to ℓ^2 [37, Theorem II.7], and the dominated convergence theorem, we find for any $\rho \in \mathcal{S}_\sigma(A)$

$$\begin{aligned} 0 \leq \limsup_{a \rightarrow \infty} \limsup_{b \rightarrow \infty} & \left[D(\rho^a \| \sigma^a) - D(\mathcal{N}^{a,b}(\rho^a) \| \mathcal{N}^{a,b}(\sigma^a)) \right. \\ & \left. + 2 \int_{\mathbb{R}} dt \beta_0(t) \log F(\rho^a, (\mathcal{R}_{\sigma^a, \mathcal{N}^{a,b}}^{\frac{t}{2}} \circ \mathcal{N}^{a,b})(\rho^a)) \right] \end{aligned} \quad (90)$$

$$\leq D(\rho \| \sigma) - D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) + 2 \int_{\mathbb{R}} dt \beta_0(t) \log F(\rho, (\mathcal{R}_{\sigma, \mathcal{N}}^{\frac{t}{2}} \circ \mathcal{N})(\rho)). \quad (91)$$

This establishes the stronger version of Theorem 2.1 in (23). We can then apply concavity of the logarithm and the fidelity to conclude Theorem 2.1.

¹Recall that by Remark 3.5, Theorem 2.1 remains valid for subnormalized states and a trace non-increasing completely positive map.

4 Proof of Theorem 2.2

For this proof we first introduce some notation. For any $H \in \mathcal{P}(A)$ where $\mathcal{P}(A)$ denotes the set of non-negative operators on A and where A is a finite-dimensional Hilbert space let $H = \sum_{k \in [d]} \lambda_k P_k$ be an eigenvalue decomposition of H , where $d \leq \dim(A)$ and $\{P_k\}_{k \in [d]}$ are mutually orthogonal projectors. For $\phi = (\phi_1, \dots, \phi_d) \in [0, 2\pi]^{\times d}$ we define the unitary

$$U_H^\phi := \sum_{k \in [d]} \exp(i\phi_k) P_k . \quad (92)$$

With the help of these unitaries we define yet another rotated Petz recovery map

$$\mathcal{T}_{\sigma, \mathcal{N}}^{\varphi, \vartheta} : X_B \mapsto U_\sigma^\vartheta \mathcal{P}_{\sigma, \mathcal{N}}(U_{\mathcal{N}(\sigma)}^\varphi X_B U_{\mathcal{N}(\sigma)}^{\varphi\dagger}) U_\sigma^{\vartheta\dagger} . \quad (93)$$

For any $\sigma \in \mathcal{P}(A)$ and $\mathcal{N} \in \text{TPCP}(A, B)$ we denote the convex hull of rotated Petz recovery maps by

$$\mathsf{T}_{\sigma, \mathcal{N}} := \text{conv} \left(\mathcal{T}_{\sigma, \mathcal{N}}^{\varphi, \vartheta} : \vartheta \in [0, 2\pi]^{\times d_1}, \varphi \in [0, 2\pi]^{\times d_2} \right) , \quad (94)$$

where $d_1 \leq \dim(A)$ and $d_2 \leq \dim(B)$. We note that the rotated Petz recovery map $\mathcal{T}_{\sigma, \mathcal{N}}^{\varphi, \vartheta}$ is trace non-increasing and completely positive. We further note that clearly $\mathsf{R}_{\sigma, \mathcal{N}} \subseteq \mathsf{T}_{\sigma, \mathcal{N}}$ with $\mathsf{R}_{\sigma, \mathcal{N}}$ defined in (16).

Step 1: Proof for a finite set of states

For any $\sigma \in \mathcal{P}(A)$ consider a finite set of density operators $\{\rho^x\}_{x \in \mathcal{X}}$ with $\rho^x \in \mathcal{S}_\sigma(A)$ for all $x \in \mathcal{X}$ and $|\mathcal{X}| < \infty$. Furthermore let ν be a probability measure on \mathcal{X} . For any channel $\mathcal{N} \in \text{TPCP}(A, B)$ we define the following non-negative operators

$$\rho_{XA} := \sum_{x \in \mathcal{X}} \nu(x) |x\rangle\langle x|_X \otimes \rho^x \quad \text{and} \quad \sigma_{XA} := \sum_{x \in \mathcal{X}} \nu(x) |x\rangle\langle x|_X \otimes \sigma = \rho_X \otimes \sigma . \quad (95)$$

This then gives

$$[(\mathcal{I}_X \otimes \mathcal{N})(\sigma_{XA})]^{-\frac{1}{2}} = [(\rho_X \otimes \mathcal{N}(\sigma))]^{-\frac{1}{2}} = \rho_X^{-\frac{1}{2}} \otimes \mathcal{N}(\sigma)^{-\frac{1}{2}} . \quad (96)$$

Since $(\mathcal{I}_X \otimes \mathcal{N})^\dagger = \mathcal{I}_X \otimes \mathcal{N}^\dagger$ and $\sigma_{XA}^{\frac{1}{2}} = (\rho_X \otimes \sigma)^{\frac{1}{2}} = \rho_X^{\frac{1}{2}} \otimes \sigma^{\frac{1}{2}}$, the Petz recovery map for σ_{XA} and $\mathcal{I}_X \otimes \mathcal{N}$ can be written as

$$\begin{aligned} Y_{XB} \mapsto \mathcal{T}_{\sigma_{XA}, \mathcal{I}_X \otimes \mathcal{N}}(Y_{XB}) &= \sigma_{XA}^{\frac{1}{2}} (\mathcal{I}_X \otimes \mathcal{N})^\dagger \left([(\mathcal{I}_X \otimes \mathcal{N})(\sigma_{XA})]^{-\frac{1}{2}} Y_{XB} [(\mathcal{I}_X \otimes \mathcal{N})(\sigma_{XA})]^{-\frac{1}{2}} \right) \sigma_{XA}^{\frac{1}{2}} \\ &= (\rho_X^{\frac{1}{2}} \otimes \sigma^{\frac{1}{2}}) (\mathcal{I}_X \otimes \mathcal{N}^\dagger) ((\rho_X^{-\frac{1}{2}} \otimes \mathcal{N}(\sigma)^{-\frac{1}{2}}) Y_{XB} (\rho_X^{-\frac{1}{2}} \otimes \mathcal{N}(\sigma)^{-\frac{1}{2}})) (\rho_X^{\frac{1}{2}} \otimes \sigma^{\frac{1}{2}}) \\ &= (\Pi_{\text{supp}(\nu)} \otimes \mathcal{T}_{\sigma, \mathcal{N}})(Y_{XB}) , \end{aligned} \quad (97)$$

where $\Pi_{\text{supp}(\nu)}$ denotes a projector onto the support of ν . Furthermore, we have that for any $\vartheta \in [0, 2\pi]^{\times d_1}$ and for any $\varphi \in [0, 2\pi]^{\times d_2}$ there exist $\tilde{\vartheta}, \bar{\vartheta} \in [0, 2\pi]^{\times d_1}$ and $\tilde{\varphi}, \bar{\varphi} \in [0, 2\pi]^{\times d_2}$ such that

$$U_{\sigma_{XA}}^\vartheta = U_{\rho_X}^{\tilde{\vartheta}} \otimes U_\sigma^{\bar{\vartheta}} \quad \text{and} \quad U_{(\Pi_{\text{supp}(\nu)} \otimes \mathcal{N})(\sigma_{XA})}^\varphi = U_{\rho_X}^{\tilde{\varphi}} \otimes U_{\mathcal{N}(\sigma)}^{\bar{\varphi}} . \quad (98)$$

Theorem 3.3 of [43] ensures that

$$\begin{aligned} &D(\rho_{XA} \| \sigma_{XA}) - D((\Pi_{\text{supp}(\nu)} \otimes \mathcal{N})(\rho_{XA}) \| (\Pi_{\text{supp}(\nu)} \otimes \mathcal{N})(\sigma_{XA})) \\ &\quad - \min \{ D_{\mathbb{M}}(\rho_{XA} \| (\mathcal{T}_{\sigma_{XA}, \mathcal{I}_X \otimes \mathcal{N}} \circ \Pi_{\text{supp}(\nu)} \otimes \mathcal{N})(\rho_{XA})) : \mathcal{T}_{\sigma_{XA}, \Pi_{\text{supp}(\nu)} \otimes \mathcal{N}} \in \mathsf{T}_{\sigma_{XA}, \Pi_{\text{supp}(\nu)} \otimes \mathcal{N}} \} \geq 0 . \end{aligned} \quad (99)$$

Combining (95), (96), (97), and (98) shows that for any $\vartheta \in [0, 2\pi]^{\times d_1}$ and for any $\varphi \in [0, 2\pi]^{\times d_2}$ there exist $\tilde{\vartheta} \in [0, 2\pi]^{\times d_1}$ and $\tilde{\varphi} \in [0, 2\pi]^{\times d_2}$ such that

$$\mathcal{T}_{\sigma, \mathcal{N}}^{\varphi, \vartheta}(\rho_{XA}) = \sum_{x \in \mathcal{X}} \nu(x) |x\rangle\langle x|_X \otimes \mathcal{T}_{\sigma, \mathcal{N}}^{\tilde{\varphi}, \tilde{\vartheta}}(\rho^x) . \quad (100)$$

This, together with a simple property of the relative and the measured relative entropy ensuring that they decompose for orthogonal states (see Lemmas B.2 and B.3 in [42]), implies that (99) can be rewritten as

$$\max_{\mathcal{T}_{\sigma, \mathcal{N}} \in \mathbb{T}_{\sigma, \mathcal{N}}} \sum_{x \in \mathcal{X}} \nu(x) \left(D(\rho^x \| \sigma) - D(\mathcal{N}(\rho^x) \| \mathcal{N}(\sigma)) - D_{\mathbb{M}}(\rho^x \| (\mathcal{T}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho^x)) \right) \geq 0 . \quad (101)$$

We note that the set $\mathbb{T}_{\sigma, \mathcal{N}}$ (defined in Equation (16)) is compact since $\text{TPCP}(B, A)$ is compact (see [42, Remark C.3]) and since the intersection of a compact set with a closed set is compact. Furthermore, the mapping $\mathbb{T}_{\sigma, \mathcal{N}} \ni \mathcal{T}_{\sigma, \mathcal{N}} \mapsto D_{\mathbb{M}}(\rho^x \| (\mathcal{T}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho^x)) \in \mathbb{R}_+$ is lower semicontinuous (see proof of Lemma 4.2). By the extreme value theorem this implies that the maximum in (101) is attained.

Since (101) is valid for any probability measure ν on \mathcal{X} we obtain

$$\min_{\nu \in \mathbb{P}(\mathcal{X})} \max_{\mathcal{T}_{\sigma, \mathcal{N}} \in \mathbb{T}_{\sigma, \mathcal{N}}} \sum_{x \in \mathcal{X}} \nu(x) \left(D(\rho^x \| \sigma) - D(\mathcal{N}(\rho^x) \| \mathcal{N}(\sigma)) - D_{\mathbb{M}}(\rho^x \| (\mathcal{T}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho^x)) \right) \geq 0 , \quad (102)$$

where $\mathbb{P}(\mathcal{X})$ denotes the set of probability measures on \mathcal{X} . We note that the minimum is attained. To see this, we first remark that the set \mathcal{X} is clearly compact and as a result, the set $\mathbb{P}(\mathcal{X})$ is weak* compact [1, Theorem 15.11]. Furthermore, the function $\mathbb{P}(\mathcal{X}) \ni \mu \mapsto \sum_{x \in \mathcal{X}} \mu(x) [D(\rho^x \| \sigma) - D(\mathcal{N}(\rho^x) \| \mathcal{N}(\sigma)) - D_{\mathbb{M}}(\rho^x \| (\mathcal{T}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho^x))]$ is continuous with respect to the weak* topology, since expectation values are continuous in the weak* topology. The extreme value theorem thus ensures that the minimum is attained.

To simplify notation, let

$$f(\nu, \mathcal{T}_{\sigma, \mathcal{N}}) := \sum_{x \in \mathcal{X}} \nu(x) \left(D(\rho^x \| \sigma) - D(\mathcal{N}(\rho^x) \| \mathcal{N}(\sigma)) - D_{\mathbb{M}}(\rho^x \| (\mathcal{T}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho^x)) \right) . \quad (103)$$

For any $\mathcal{T}_{\sigma, \mathcal{N}} \in \mathbb{T}_{\sigma, \mathcal{N}}$, the function $\mathbb{P}(\mathcal{X}) \ni \nu \mapsto f(\nu, \mathcal{T}_{\sigma, \mathcal{N}}) \in \mathbb{R}$ is linear and bounded (since $\rho^x \in \mathcal{S}_{\sigma}(A)$ for all $x \in \mathcal{X}$) and hence continuous. Furthermore, for any $\nu \in \mathbb{P}(\mathcal{X})$ the function $\mathbb{T}_{\sigma, \mathcal{N}} \ni \mathcal{T}_{\sigma, \mathcal{N}} \mapsto f(\nu, \mathcal{T}_{\sigma, \mathcal{N}}) \in \mathbb{R}$ is concave [42, Lemma B.4] and upper semicontinuous (as shown in the proof of Lemma 4.2). The sets $\mathbb{P}(\mathcal{X})$ and $\mathbb{T}_{\sigma, \mathcal{N}}$ are both convex and as discussed above $\mathbb{T}_{\sigma, \mathcal{N}}$ is compact. As a result, Sion's minimax theorem [41, Corollary 3.3] implies that

$$\max_{\mathcal{T}_{\sigma, \mathcal{N}} \in \mathbb{T}_{\sigma, \mathcal{N}}} \min_{\nu \in \mathbb{P}(\mathcal{X})} \sum_{x \in \mathcal{X}} \nu(x) \left(D(\rho^x \| \sigma) - D(\mathcal{N}(\rho^x) \| \mathcal{N}(\sigma)) - D_{\mathbb{M}}(\rho^x \| (\mathcal{T}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho^x)) \right) \geq 0 . \quad (104)$$

This shows that there exists a recovery map $\mathcal{T}_{\sigma, \mathcal{N}} \in \mathbb{T}_{\sigma, \mathcal{N}}$ that satisfies

$$D(\rho^x \| \sigma) - D(\mathcal{N}(\rho^x) \| \mathcal{N}(\sigma)) \geq D_{\mathbb{M}}(\rho^x \| (\mathcal{T}_{\sigma, \mathcal{N}} \circ \mathcal{N})(\rho^x)) \quad \text{for all } x \in \mathcal{X} . \quad (105)$$

Since $\mathcal{R}_{\sigma, \mathcal{N}} \subseteq \mathbb{T}_{\sigma, \mathcal{N}}$, this proves the statement of Theorem 2.2 for a finite set of states $\{\rho^x\}_{x \in \mathcal{X}}$.

Step 2: Extensions to an infinite set of states

We now show how to extend the statement proven in Step 1 for a finite set of states to the set of all states. Let $\mathcal{S} := \mathcal{S}(A)$ denote the set of density operators on A . Furthermore, for $\mathcal{T} \in \text{TPCP}(B, A)$, $\sigma \in \mathcal{P}(A)$, and $\mathcal{N} \in \text{TPCP}(A, B)$ we define the function family

$$\begin{aligned} \Delta_{\mathcal{T}, \sigma, \mathcal{N}} : \mathcal{S}(A) &\rightarrow \mathbb{R} \cup \{-\infty\} \\ \rho &\mapsto D(\rho \| \sigma) - D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) - D_{\mathbb{M}}(\rho \| (\mathcal{T} \circ \mathcal{N})(\rho)) . \end{aligned} \quad (106)$$

We start by proving two regularity properties of the Δ -function family defined above.

Lemma 4.1. *For any $\mathcal{T} \in \text{TPCP}(B, A)$, any $\sigma \in \mathcal{P}(A)$, and any $\mathcal{N} \in \text{TPCP}(A, B)$, the function $S_\sigma(A) \ni \rho \mapsto \Delta_{\mathcal{T}, \sigma, \mathcal{N}}(\rho) \in \mathbb{R}$ defined in (106) is upper semicontinuous.*

Proof. For any fixed $\sigma \in \mathcal{P}(A)$, the function $S_\sigma(A) \ni \rho \mapsto D(\rho \| \sigma) = \text{tr}(\rho(\log \rho - \log \sigma)) = -H(\rho) - \text{tr}(\rho \log \sigma)$ is continuous. This follows by Fannes' inequality [10] that shows that $S(A) \ni \rho \mapsto H(\rho) \in \mathbb{R}$ is continuous. Furthermore the function $S_\sigma(A) \ni \rho \mapsto -\text{tr}(\rho \log \sigma) \in \mathbb{R}$ is bounded and linear and thus continuous. Since $S(A) \ni \rho \mapsto \mathcal{N}(\rho)$ is linear and bounded it is continuous and thus the function $S_\sigma(A) \ni \rho \mapsto D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma))$ is also continuous. In addition, the function $S_\sigma(A) \ni \rho \mapsto D_{\mathbb{M}}(\rho \| (\mathcal{T} \circ \mathcal{N})(\rho))$ is lower semicontinuous [42, Lemma 3.2] which then proves the assertion. \square

Lemma 4.2. *For any $\sigma \in S(A)$, any $\rho \in S_\sigma(A)$, and any $\mathcal{N} \in \text{TPCP}(A, B)$, the function $\text{TPCP}(B, A) \ni \mathcal{T} \mapsto \Delta_{\mathcal{T}, \sigma, \mathcal{N}}(\rho) \in \mathbb{R}$ defined in (106) is upper semicontinuous.*

Proof. The assertion follows from the fact that the mapping $\text{TPCP}(B, A) \ni \mathcal{T} \mapsto D_{\mathbb{M}}(\rho \| (\mathcal{T} \circ \mathcal{N})(\rho)) \in \mathbb{R}_+$ is lower semicontinuous. To see this, we first note that by definition we have $D_{\mathbb{M}}(\rho \| (\mathcal{T} \circ \mathcal{N})(\rho)) = \sup_{\mathcal{M} \in \mathbb{M}} D(\mathcal{M}(\rho) \| \mathcal{M}((\mathcal{T} \circ \mathcal{N})(\rho)))$. Since \mathcal{T} , \mathcal{N} , and \mathcal{M} are linear bounded operators and hence continuous we find that $\mathcal{T} \mapsto D(\mathcal{M}(\rho^x) \| \mathcal{M}((\mathcal{T} \circ \mathcal{N})(\rho^x)))$ is continuous as the logarithm $\mathbb{R}^+ \ni x \mapsto \log x \in \mathbb{R}$ is continuous. As the supremum of continuous functions is lower semicontinuous [6, Chapter IV, Section 6.2, Theorem 4], the assertion follows. \square

Since S is compact (see, e.g., [42, Lemma C.1]), it follows that for any $\varepsilon > 0$ there exists a finite set S^ε of density operators on A such that any $\rho \in S$ is ε -close to an element of S^ε . We assume without loss of generality that $S^{\varepsilon'} \subset S^\varepsilon$ for $\varepsilon' \geq \varepsilon$. Let $\mathcal{T}^\varepsilon \in \text{TPCP}(B, A)$ be such that $\inf_{\rho \in S^\varepsilon} \Delta_{\mathcal{T}^\varepsilon, \sigma, \mathcal{N}}(\rho) \geq 0$ whose existence has been established in Step 1. The set $\text{TPCP}(B, A)$ is compact [42, Remark C.3] and the intersection of a compact set with a closed set is compact. Since $\text{TPCP}(B, A)$ is compact, there exists a decreasing sequence $\{\varepsilon_n\}_{n \in \mathbb{N}}$ and $\tilde{\mathcal{T}} \in \text{TPCP}(B, A)$ such that

$$\lim_{n \rightarrow \infty} \varepsilon_n = 0 \quad \text{and} \quad \tilde{\mathcal{T}} = \lim_{n \rightarrow \infty} \mathcal{T}^{\varepsilon_n}. \quad (107)$$

Thus for all $n \in \mathbb{N}$ we find by Lemma 4.2

$$\begin{aligned} \inf_{\rho \in S^{\varepsilon_n}} \Delta_{\tilde{\mathcal{T}}, \sigma, \mathcal{N}}(\rho) &\geq \inf_{\rho \in S^{\varepsilon_n}} \limsup_{m \rightarrow \infty} \Delta_{\mathcal{T}^{\varepsilon_m}, \sigma, \mathcal{N}}(\rho) \geq \limsup_{m \rightarrow \infty} \inf_{\rho \in S^{\varepsilon_n}} \Delta_{\mathcal{T}^{\varepsilon_m}, \sigma, \mathcal{N}}(\rho) \\ &\geq \limsup_{m \rightarrow \infty} \inf_{\rho \in S^{\varepsilon_m}} \Delta_{\mathcal{T}^{\varepsilon_m}, \sigma, \mathcal{N}}(\rho) \geq 0, \end{aligned} \quad (108)$$

where the final step follows by Step 1. For any $\rho \in S$ let $\rho_n \in S^{\varepsilon_n}$ be such that $\lim_{n \rightarrow \infty} \rho_n = \rho$. (By the definition of S^{ε_n} , it follows that such a sequence $\{\rho_n\}_{n \in \mathbb{N}}$ with $\rho_n \in S^{\varepsilon_n}$ always exists.) By Lemma 4.1 and (108) we find for any $\rho \in S$

$$\Delta_{\tilde{\mathcal{T}}, \sigma, \mathcal{N}}(\rho) = \Delta_{\tilde{\mathcal{T}}, \sigma, \mathcal{N}}(\lim_{n \rightarrow \infty} \rho_n) \geq \limsup_{n \rightarrow \infty} \Delta_{\tilde{\mathcal{T}}, \sigma, \mathcal{N}}(\rho_n) \geq \limsup_{n \rightarrow \infty} \inf_{\rho \in S^{\varepsilon_n}} \Delta_{\tilde{\mathcal{T}}, \sigma, \mathcal{N}}(\rho) \geq 0. \quad (109)$$

Since $R_{\sigma, \mathcal{N}} \subseteq \text{TPCP}(B, A)$, this proves the first inequality in (21).

The second inequality in (21) is a consequence of the monotonicity of the Rényi divergence in the order parameter [8] and of the fact that for any two states there exists an optimal measurement that does not increase their fidelity [14, Section 3.3]. This thus concludes the proof of Theorem 2.2 for finite-dimensional Hilbert spaces.

5 Universal refinements of other entropy inequalities

It is well known (see e.g. [38]) that the monotonicity of the relative entropy under trace-preserving completely positive maps is closely related to (i) strong subadditivity, (ii) concavity of the conditional entropy, and (iii) joint convexity of the relative entropy. Based on this relation, Theorem 2.2 can be used to derive universal remainder terms for the statements (i)-(iii). We note that universal means that

the recovery map that occurs in the remainder term does not depend on all possible parameters. Within this section we assume that A , B , and C are finite-dimensional Hilbert spaces. For $n \in \mathbb{N}$, the n -simplex is denoted by $\Delta_n := \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x_i = 1\}$. The following three corollaries consist of two statements each. Statement 1 is derived from Theorem 2.1 which has the advantage that the remainder term is explicit (and hence also universal). However, since the lower bound of Theorem 2.1 is in form of a logarithm of a fidelity we cannot fully simplify the remainder terms (e.g., the convex sum appears inside the logarithm). This is circumvented in Statement 2 (which is derived from Theorem 2.2)² at the cost that the remainder term is not explicit, but still universal.

Applying Theorems 2.1 and 2.2 for $\sigma = \text{id}_A \otimes \rho_{BC} \in \mathcal{P}(A \otimes B \otimes C)$, $\rho = \rho_{ABC} \in \mathcal{S}_\sigma(A \otimes B \otimes C)$, and $\mathcal{N}_{ABC \rightarrow AB} = \mathcal{I}_{AB} \otimes \text{tr}_C$ implies a strengthened version of the result that has been established in [42, Theorem 2.1, Corollary 2.4, and Remark 2.5].

Corollary 5.1 (Strong subadditivity).

1. For any density operator $\rho_{ABC} \in \mathcal{S}(A \otimes B \otimes C)$ we have

$$I(A : C|B)_\rho \geq -2 \log F(\rho_{ABC}, \mathcal{R}_{B \rightarrow BC}(\rho_{AB})) , \quad (110)$$

where $\mathcal{R}_{B \rightarrow BC}(\cdot) = \int_{\mathbb{R}} dt \beta_0(t) \mathcal{R}_{\rho_{BC}, \text{tr}_C}^t(\cdot)$ as defined in Theorem 2.1.

2. For any density operator $\rho_{BC} \in \mathcal{S}(B \otimes C)$ there exists a recovery map $\mathcal{R}_{B \rightarrow BC}$ inside the set $\mathcal{R}_{\rho_{BC}, \text{tr}_C}$ defined in (16), such that for any extension ρ_{ABC} on $A \otimes B \otimes C$ we have

$$I(A : C|B)_\rho \geq D_{\mathbb{M}}(\rho_{ABC} \| \mathcal{R}_{B \rightarrow BC}(\rho_{AB})) \geq -2 \log F(\rho_{ABC}, \mathcal{R}_{B \rightarrow BC}(\rho_{AB})) . \quad (111)$$

From Corollary 5.1 we can deduce a universal remainder term for the concavity of the conditional entropy. We note that a similar remainder term has been conjectured in [5]. Furthermore, a remainder term that however is neither universal nor explicit has been proven in [45, Corollary 12].

Corollary 5.2 (Concavity of conditional entropy).

1. For any $\rho_{AB} \in \mathcal{S}(A \otimes B)$ we have

$$H(A|B)_\rho - \sum_{x \in \mathcal{X}} \nu(x) H(A|B)_{\rho^x} \geq -2 \log \sum_{x \in \mathcal{X}} \nu(x) F(\rho_{AB}^x, \mathcal{R}_{B \rightarrow AB}(\rho_B^x)) , \quad (112)$$

for all ensembles $\{\nu(x), \rho_{AB}^x\}_{x \in \mathcal{X}}$ with $\nu \in \Delta_{\mathcal{X}}$ and $\rho_{AB}^x \in \mathcal{S}(A \otimes B)$ such that $\rho_{AB} = \sum_{x \in \mathcal{X}} \nu(x) \rho_{AB}^x$ where $\mathcal{R}_{B \rightarrow AB}(\cdot) = \int_{\mathbb{R}} dt \beta_0(t) \mathcal{R}_{\rho_{AB}, \text{tr}_A}^t(\cdot)$ as defined in Theorem 2.1.

2. For any $\rho_{AB} \in \mathcal{S}(A \otimes B)$ there exists recovery map $\mathcal{R}_{B \rightarrow AB} \in \mathcal{R}_{\rho_{AB}, \text{tr}_A}$ with $\mathcal{R}_{\rho_{AB}, \text{tr}_A}$ defined in (16), such that

$$\begin{aligned} H(A|B)_\rho - \sum_{x \in \mathcal{X}} \nu(x) H(A|B)_{\rho^x} &\geq \sum_{x \in \mathcal{X}} \nu(x) D_{\mathbb{M}}(\rho_{AB}^x \| \mathcal{R}_{B \rightarrow AB}(\rho_B^x)) \\ &\geq -2 \sum_{x \in \mathcal{X}} \nu(x) \log F(\rho_{AB}^x, \mathcal{R}_{B \rightarrow AB}(\rho_B^x)) , \end{aligned} \quad (113)$$

for all ensembles $\{\nu(x), \rho_{AB}^x\}_{x \in \mathcal{X}}$ with $\nu \in \Delta_{\mathcal{X}}$ and $\rho_{AB}^x \in \mathcal{S}(A \otimes B)$ such that $\rho_{AB} = \sum_{x \in \mathcal{X}} \nu(x) \rho_{AB}^x$.

Proof. We will provide a proof for Statement 2. Statement 1 follows by exactly the same reasoning. Let us consider the following classical-quantum state

$$\phi_{XAB} = \sum_{x \in \mathcal{X}} \nu(x) |x\rangle\langle x|_X \otimes \rho_{AB}^x . \quad (114)$$

²Recall that the lower bound in Theorem 2.2 is in the form of a measured relative entropy.

By Corollary 5.1 and a simple fact ensuring that the measured relative entropy for orthogonal input states decomposes (see Lemma B.3 in [42]), we find

$$\begin{aligned} H(A|B)_\rho - \sum_{x \in \mathcal{X}} \nu(x) H(A|B)_{\rho^x} &= H(A|B)_\phi - H(A|BX)_\phi = I(X : A|B)_\phi \\ &\geq D_{\mathbb{M}}(\phi_{XAB} \| \mathcal{R}_{B \rightarrow AB}(\phi_{XB})) = \sum_{x \in \mathcal{X}} \nu(x) D_{\mathbb{M}}(\rho_{AB}^x \| \mathcal{R}_{B \rightarrow AB}(\rho_B^x)) . \end{aligned} \quad (115)$$

We note that by Corollary 5.1 the recovery map $\mathcal{R}_{B \rightarrow AB}$ is universal in the sense that it only depends on ρ_{AB} . The second inequality in (113) is a consequence of the monotonicity of the Rényi divergence in the order parameter [8] and of the fact that for any two states there exists an optimal measurement that does not increase their fidelity [14, Section 3.3]. \square

Theorems 2.1 and 2.2 imply a universal remainder term for the joint convexity of the relative entropy. We note that a similar remainder term has been conjectured in [39]. Corollary 13 in [45] proves such a remainder term that however is neither universal nor explicit.

Corollary 5.3 (Joint convexity of relative entropy).

1. For any $\nu \in \Delta_{\mathcal{X}}$, $\{\sigma^x\}_{x \in \mathcal{X}}$ with $\sigma^x \in \mathcal{P}(A)$ for all $x \in \mathcal{X}$, $\sigma_{XA} = \sum_{x \in \mathcal{X}} \nu(x) |x\rangle\langle x|_X \otimes \sigma^x$, any $\rho_x \in \mathcal{S}_{\sigma^x}(A)$ and $\rho = \sum_{x \in \mathcal{X}} \nu(x) \rho^x$ we have

$$\sum_{x \in \mathcal{X}} \nu(x) D(\rho^x \| \sigma^x) - D(\rho \| \sigma) \geq -2 \log \sum_{x \in \mathcal{S}} \nu(x) F(\rho^x, \mathcal{R}_{\sigma_{XA}, \text{tr}_X}(\rho)) , \quad (116)$$

where $\mathcal{R}_{\sigma_{XA}, \text{tr}_X}(\cdot) = \int_{\mathbb{R}} dt \beta_0(t) \mathcal{R}_{\sigma_{XA}, \text{tr}_X}^t(\cdot)$ as defined in Theorem 2.1.

2. For any $\nu \in \Delta_{\mathcal{X}}$, $\{\sigma^x\}_{x \in \mathcal{X}}$ with $\sigma^x \in \mathcal{P}(A)$ for all $x \in \mathcal{X}$, and $\sigma_{XA} = \sum_{x \in \mathcal{X}} \nu(x) |x\rangle\langle x|_X \otimes \sigma^x$ there exists a recovery map $\mathcal{R} \in \mathcal{R}_{\sigma_{XA}, \text{tr}_X}$ such that

$$\sum_{x \in \mathcal{X}} \nu(x) D(\rho^x \| \sigma^x) - D(\rho \| \sigma) \geq \sum_{x \in \mathcal{X}} \nu(x) D_{\mathbb{M}}(\rho^x \| \mathcal{R}(\rho)) \geq -2 \sum_{x \in \mathcal{X}} \nu(x) \log F(\rho^x, \mathcal{R}(\rho)) \quad (117)$$

for all $\{\rho^x\}_{x \in \mathcal{X}}$ with $\rho_x \in \mathcal{S}_{\sigma^x}(A)$ and $\rho = \sum_{x \in \mathcal{X}} \nu(x) \rho^x$.

Proof. We will provide a proof for Statement 2. Statement 1 follows by exactly the same reasoning. Let us consider the states

$$\rho_{XA} := \sum_{x \in \mathcal{X}} \nu(x) |x\rangle\langle x|_X \otimes \rho^x \quad \text{and} \quad \rho = \rho_A = \sum_{x \in \mathcal{X}} \nu(x) \rho^x . \quad (118)$$

Since the relative entropy decomposes for orthogonal input states (see Lemma B.2 in [42]), Theorem 2.2 implies that there exists a recovery map $\mathcal{R} \in \mathcal{R}_{\sigma_{XA}, \text{tr}_X}$ such that

$$\begin{aligned} \sum_{x \in \mathcal{X}} \nu(x) D(\rho^x \| \sigma^x) - D(\rho \| \sigma) &= D(\rho_{XA} \| \sigma_{XA}) - D(\rho_A \| \sigma_A) \\ &\geq D_{\mathbb{M}}(\rho_{XA} \| \mathcal{R}(\rho)) = \sum_{x \in \mathcal{X}} \nu(x) D_{\mathbb{M}}(\rho^x \| \mathcal{R}(\rho)) . \end{aligned} \quad (119)$$

The second inequality in (117) is a consequence of the monotonicity of quantum Rényi divergence in the order parameter [32] and of the fact that for any two states there exists an optimal measurement that does not increase their fidelity [14, Section 3.3]. \square

6 Approximate quantum error correction

Theorem 2.1 allows for establishing an alternate, information-theoretic set of conditions for approximate quantum error correction. Recall that in quantum error correction the main goal is to protect a set of states from the action of a noisy quantum channel, by encoding this set into a subspace of a given Hilbert space. The quantum error correction conditions given in the seminal works [4, 24] are necessary and sufficient for perfect quantum error correction. Although these works were essential for gaining an understanding of quantum error correction, the conditions given represent a strong idealization, since we can never hope in practice to have perfect quantum error correction. So this realization motivated some follow-up works on approximate quantum error correction (see [3, 30] and references therein), in which one seeks to determine conditions under which approximate recovery from the action of a given noisy channel is possible. We mention that some of these works made essential use of the Petz recovery map.

In [17], information-theoretic conditions for perfect error correction were given in terms of the quantum relative entropy. We recall their result: Let A be a Hilbert space, let C be a subspace of A (called the “codespace”), and let Π denote the projection onto C . Then

$$\{\forall \rho \in S(C) : D(\rho \| \Pi) = D(\mathcal{N}(\rho) \| \mathcal{N}(\Pi))\} \iff \{\forall \rho \in S(C) : \rho = (\mathcal{P}_{\Pi, \mathcal{N}} \circ \mathcal{N})(\rho)\} , \quad (120)$$

where $\mathcal{P}_{\Pi, \mathcal{N}}(\cdot)$ is the Petz recovery map defined in (4). *So perfect error correction is possible if and only if for all $\rho \in S(C)$ the pairwise distinguishability of ρ with Π does not decrease under the action of the noisy channel \mathcal{N} .*

Given the statement in (120) and the prior work on approximate quantum error correction, it is natural to wonder whether a robust version of (120) exists. Indeed, Theorem 2.1 implies such a statement, establishing necessary and sufficient information-theoretic conditions for approximate quantum error correction. Loosely speaking, we can now say that *approximate error correction is possible if and only if the pairwise distinguishability of ρ with Π does not decrease by too much under the action of the noisy channel \mathcal{N} for all $\rho \in S(C)$* . The corollary below is a consequence of Theorem 2.1 and some other known facts.

Corollary 6.1. *Let $\varepsilon \in [0, 1]$, A and B be a finite-dimensional Hilbert spaces, C be a subspace of A (called the “codespace”), $\mathcal{N} \in \text{TPCP}(A, B)$ a quantum channel, and let Π denote the projection onto C . If for all $\rho \in S(C)$ we have*

$$D(\rho \| \Pi) - D(\mathcal{N}(\rho) \| \mathcal{N}(\Pi)) \leq \varepsilon , \quad (121)$$

then it is possible to approximately recover every state $\rho \in S(C)$, in the sense that we have for all $\rho \in S(C)$

$$F(\rho, (\mathcal{R}_{\Pi, \mathcal{N}} \circ \mathcal{N})(\rho)) \geq 1 - \frac{1}{2}\varepsilon . \quad (122)$$

Conversely, if (122) holds for $\varepsilon \in [0, 1]$, then we have for all $\rho \in S(C)$

$$D(\rho \| \Pi) - D(\mathcal{N}(\rho) \| \mathcal{N}(\Pi)) \leq \sqrt{\varepsilon} \log(\dim C) + h_2(\sqrt{\varepsilon}) , \quad (123)$$

where $h_2(p) := -p \log p - (1-p) \log(1-p)$ is the binary entropy, with the property that $\lim_{p \searrow 0} h_2(p) = 0$.

Proof. The first statement is a direct consequence of Theorem 2.1, found by setting $\rho = \rho$, $\sigma = \Pi$, and $\mathcal{N} = \mathcal{N}$ and then applying the inequality $-\log(x) \geq 1 - x$, which holds for $x \in [0, 1]$.

To prove the second statement, suppose that (122) holds. By the Fuchs-van-de-Graaf inequality [13], (122) implies that for all $\rho \in S(C)$ we have

$$\frac{1}{2} \|\rho - (\mathcal{R}_{\Pi, \mathcal{N}} \circ \mathcal{N})(\rho)\|_1 \leq \sqrt{\varepsilon} . \quad (124)$$

Then consider the following chain of inequalities, which holds for all $\rho \in \mathcal{S}(C)$:

$$D(\rho\|\Pi) - D(\mathcal{N}(\rho)\|\mathcal{N}(\Pi)) \leq D(\rho\|\Pi) - D((\mathcal{R}_{\Pi,\mathcal{N}} \circ \mathcal{N})(\rho)\|(\mathcal{R}_{\Pi,\mathcal{N}} \circ \mathcal{N})(\Pi)) \quad (125)$$

$$= D(\rho\|\Pi) - D((\mathcal{R}_{\Pi,\mathcal{N}} \circ \mathcal{N})(\rho)\|\Pi) \quad (126)$$

$$= -H(\rho) + H((\mathcal{R}_{\Pi,\mathcal{N}} \circ \mathcal{N})(\rho)) + \text{tr}([(\mathcal{R}_{\Pi,\mathcal{N}} \circ \mathcal{N})(\rho) - \rho] \log \Pi) \quad (127)$$

$$\leq \sqrt{\varepsilon} \log \dim(C) + h_2(\sqrt{\varepsilon}). \quad (128)$$

The first inequality is a consequence of monotonicity of quantum relative entropy with respect to the recovery map $\mathcal{R}_{\Pi,\mathcal{N}}$. The first equality follows because $(\mathcal{R}_{\Pi,\mathcal{N}} \circ \mathcal{N})(\Pi) = \Pi$. The second equality is a rewriting, and the last inequality follows from the Fannes-Audenaert inequality [2] (see also [46, Lemma 1]) and the facts that $\text{supp}(\rho) \subseteq \text{supp}(\Pi)$, $\text{supp}((\mathcal{R}_{\Pi,\mathcal{N}} \circ \mathcal{N})(\rho)) \subseteq \text{supp}(\Pi)$, and $\log \Pi = 0$ on the support of Π . \square

7 Conclusion

In this work, we showed that for any non-negative operator σ and for any channel \mathcal{N} there exists an *explicit* and *universal* recovery map $\mathcal{R}_{\sigma,\mathcal{N}}$ and

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq -2 \log F(\rho, (\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)) \quad (129)$$

for all density operators ρ such that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$. A consequence of its universality is that the recovery map $\mathcal{R}_{\sigma,\mathcal{N}}$ satisfies $(\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\sigma) = \sigma$. The present work thus constitutes an improvement over [45] since the recovery map is constructed explicitly and is independent of ρ . Knowing the explicit structure of the recovery map that satisfies (129) can be helpful in various scenarios. For example if this is to be implemented by experimentalists or in an actual quantum computer it is helpful to have an explicit recovery map. We showed that the inequality (129) implies universal and explicit remainder terms for other entropy inequalities such as strong subadditivity, concavity of the entropy, and the joint convexity of the relative entropy. Furthermore, it can be useful in the context of quantum sufficient statistics and approximate quantum error correction.

Two questions that remain unanswered are, first, if (129) can be tightened in the sense that its right-hand side can be replaced by a relative entropy term of the form $D(\rho\|\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)$. Secondly, it is also unknown if a bound in terms of a measured relative entropy (in the form of (11)) remains valid for an *explicit* recovery map — recall that Theorem 2.2 proves the existence of a universal recovery map that satisfies (11) but this recovery map is not explicit. Theorem 2.1, which shows an explicit recovery map that satisfies (129), could therefore provide a new approach to attack these two open questions.

Appendix

A Proof of Lemma 3.3

We begin by recalling Hirschman's strengthening [18] of the Hadamard three-line theorem (see, e.g., [16, Lemma 1.3.8]):

Lemma A.1. *Let $S := \{z \in \mathbb{C} : 0 \leq \text{Re}\{z\} \leq 1\}$ and let $g(z)$ be holomorphic on the interior of S and continuous on the boundary, such that*

$$\sup_{z \in S} \exp\{-a |\text{Im } z|\} \log |g(z)| \leq A < \infty, \quad (130)$$

for some fixed A and $a < \pi$. Then for $\theta \in (0, 1)$, the following bound holds

$$\log |g(\theta)| \leq \int_{\mathbb{R}} dt \left(\alpha_{\theta}(t) \log(|g(it)|^{1-\theta}) + \beta_{\theta}(t) \log(|g(1+it)|^{\theta}) \right), \quad (131)$$

where $\alpha_{\theta}(t)$ and $\beta_{\theta}(t)$ are defined in (30).

We can now prove Lemma 3.3.

Proof of Lemma 3.3. The proof of this theorem is well known, but we provide it for completeness. For fixed $\theta \in (0, 1)$, let q_θ be the Hölder conjugate of p_θ , defined by

$$\frac{1}{p_\theta} + \frac{1}{q_\theta} = 1. \quad (132)$$

Similarly, let q_0 and q_1 be Hölder conjugates of p_0 and p_1 , respectively. We can find an operator X such that

$$\|X\|_{q_\theta} = 1 \quad \text{and} \quad \text{tr}(XG(\theta)) = \|G(\theta)\|_{p_\theta}. \quad (133)$$

We can write the singular value decomposition for X in the form $X = UD^{1/q_\theta}V$ (implying $\text{tr}(D) = 1$). For $z \in S$, define

$$X(z) := UD^{\frac{1-z}{q_0} + \frac{z}{q_1}}V. \quad (134)$$

As a consequence, $X(z)$ is holomorphic on the interior of S and continuous on the boundary. Also, observe that $X(\theta) = X$. Then the following function satisfies the requirements needed to apply Lemma A.1:

$$g(z) := \text{tr}(X(z)G(z)). \quad (135)$$

Indeed, we have that

$$\log \|G(\theta)\|_{p_\theta} = \log |g(\theta)| \leq \int_{\mathbb{R}} dt \left(\alpha_\theta(t) \log(|g(it)|^{1-\theta}) + \beta_\theta(t) \log(|g(1+it)|^\theta) \right). \quad (136)$$

Now, from applying Hölder's inequality and the facts that $\|X(it)\|_{q_0} = 1 = \|X(1+it)\|_{q_1}$, we find that

$$|g(it)| = |\text{tr}(X(it)G(it))| \leq \|X(it)\|_{q_0} \|G(it)\|_{p_0} = \|G(it)\|_{p_0} \quad (137)$$

and

$$|g(1+it)| = |\text{tr}(X(1+it)G(1+it))| \leq \|X(1+it)\|_{q_1} \|G(1+it)\|_{p_1} = \|G(1+it)\|_{p_1}. \quad (138)$$

Bounding (136) from above using these inequalities then gives (29). \square

Acknowledgments

We thank Omar Fawzi, Rupert Frank, Elliott Lieb, Volkher Scholz and Marco Tomamichel for helpful discussions. MJ's work was supported by NSF DMS 1501103 and NSF-DMS 1201886. RR and DS acknowledge support by the European Research Council (ERC) via grant No. 258932, by the Swiss National Science Foundation (SNSF) via the National Centre of Competence in Research "QSIT", and by the European Commission via the project "RAQUEL". MMW acknowledges support from the NSF under Award No. CCF-1350397 and the DARPA Quiness Program through US Army Research Office award W31P4Q-12-1-0019. He is also grateful to co-authors MJ, RR, and AW for hospitality and support during research visits in summer 2015. AWs work was supported by the EU (STREP RAQUEL), the ERC (AdG IRQUAT), the Spanish MINECO (grant FIS2013-40627-P) with the support of FEDER funds, as well as by the Generalitat de Catalunya CIRIT, project 2014-SGR-966.

References

- [1] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis*. Springer, 2007.
- [2] K. M. R. Audenaert. A sharp continuity estimate for the von Neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127, 2007.

- [3] H. Barnum and E. Knill. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics*, 43(5):2097–2106, May 2002. [arXiv:quant-ph/0004088](#).
- [4] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824–3851, November 1996.
- [5] M. Berta, M. Lemm, and M. M. Wilde. Monotonicity of quantum relative entropy and recoverability. *Quantum Information and Computation*, 15(15 & 16):1333–1354, November 2015.
- [6] N. Bourbaki. *Elements of Mathematics: General Topology*. Hermann, Éditeurs des Sciences et des Arts, 1966.
- [7] F. G. S. L. Brandão, A. W. Harrow, J. Oppenheim, and S. Strelchuk. Quantum conditional mutual information, reconstructed states, and state redistribution. *Physical Review Letters*, 115(5):050501, July 2015.
- [8] I. Csiszár. Generalized cutoff rates and Rényi’s information measures. *IEEE Transactions on Information Theory*, 41(1):26–34, Jan 1995.
- [9] I. Devetak and J. Yard. Exact cost of redistributing multipartite quantum states. *Physical Review Letters*, 100(23):230501, June 2008.
- [10] M. Fannes. A continuity property of the entropy density for spin lattice systems. *Communications in Mathematical Physics*, 31(4):291–294, 1973.
- [11] O. Fawzi and R. Renner. Quantum conditional mutual information and approximate Markov chains. *Communications in Mathematical Physics*, 340(2):575–611, 2015.
- [12] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London Series A*, 222:309–368, 1922.
- [13] C. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, May 1999.
- [14] C. A. Fuchs. Distinguishability and accessible information in quantum theory. *PhD Thesis, University of New Mexico*, 1996. [arXiv:quant-ph/9601020](#).
- [15] F. Furrer, J. Åberg, and R. Renner. Min- and max-entropy in infinite dimensions. *Communications in Mathematical Physics*, 306(1):165–186, 2011.
- [16] L. Grafakos. *Classical Fourier Analysis*. Springer, second edition, 2008.
- [17] F. Hiai, M. Mosonyi, D. Petz, and C. Beny. Quantum f -divergences and error correction. *Reviews in Mathematical Physics*, 23(7):691–747, August 2011. [arXiv:1008.2529](#).
- [18] I. I. Hirschman. A convexity theorem for certain groups of transformations. *Journal d’Analyse Mathématique*, 2(2):209–218, December 1952.
- [19] A. S. Holevo. *Quantum Systems, Channels, Information*. De Gruyter Studies in Mathematical Physics 16, 2012.
- [20] A. Jencová. Randomization theorems for quantum channels, 2015. [arXiv:1404.3900](#).
- [21] A. Jencová and D. Petz. Sufficiency in quantum statistical inference. *Communications in Mathematical Physics*, 263(1):259–276, 2006.
- [22] A. Jencová and D. Petz. Sufficiency in quantum statistical inference: A survey with examples. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 09(03):331–351, 2006.

- [23] R. V. Kadison and J. R. Ringrose. *Fundamentals of the Theory of Operator Algebras*, volume I: Elementary Theory. Academic Press, 1983.
- [24] E. Knill and R. Laflamme. Theory of quantum error-correcting codes. *Physical Review A*, 55(2):900–911, February 1997.
- [25] K. Li and A. Winter. Squashed entanglement, k-extendibility, quantum Markov chains, and recovery maps, 2014. [arXiv:1410.4184](https://arxiv.org/abs/1410.4184).
- [26] E. H. Lieb and M. B. Ruskai. A fundamental property of quantum-mechanical entropy. *Phys. Rev. Lett.*, 30:434–436, Mar 1973.
- [27] E. H. Lieb and M. B. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14(12):1938–1941, 1973.
- [28] G. Lindblad. Expectations and entropy inequalities for finite quantum systems. *Communications in Mathematical Physics*, 39(2):111–119, 1974.
- [29] G. Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40(2):147–151, 1975.
- [30] P. Mandayam and H. K. Ng. Towards a unified framework for approximate quantum error correction. *Physical Review A*, 86(1):012335, July 2012. [arXiv:1202.5139](https://arxiv.org/abs/1202.5139).
- [31] M. Mosonyi and D. Petz. Structure of sufficient quantum coarse-grainings. *Letters in Mathematical Physics*, 68(1):19–30, 2004.
- [32] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel. On quantum Rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, 2013.
- [33] M. Ohya and D. Petz. *Quantum Entropy and Its Use*. Springer-Verlag, 1993.
- [34] D. Petz. Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Communications in Mathematical Physics*, 105(1):123–131, 1986.
- [35] D. Petz. Sufficiency of channels over von Neumann algebras. *The Quarterly Journal of Mathematics*, 39(1):97–108, 1988.
- [36] D. Petz. Monotonicity of quantum relative entropy revisited. *Reviews in Mathematical Physics*, 15(01):79–91, 2003.
- [37] M. Reed and B. Simon. *Functional Analysis*. Elsevier, Academic Press, 1980.
- [38] M. B. Ruskai. Inequalities for quantum entropy: A review with conditions for equality. *Journal of Mathematical Physics*, 43(9):4358–4375, 2002.
- [39] K. P. Seshadreesan, M. Berta, and M. M. Wilde. Rényi squashed entanglement, discord, and relative entropy differences. *Journal of Physics A: Mathematical and Theoretical*, 48(39):395303, 2015.
- [40] M. E. Shirokov and A. S. Holevo. On approximation of infinite-dimensional quantum channels. *Problems of Information Transmission*, 44(2):3–22, 2008. [arXiv:0711.2245](https://arxiv.org/abs/0711.2245).
- [41] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8:171, 1958.
- [42] D. Sutter, O. Fawzi, and R. Renner. Universal recovery map for approximate Markov chains, 2015. accepted for publication at Proceedings of the Royal Society of London A; available at [arXiv:1504.07251](https://arxiv.org/abs/1504.07251).

- [43] D. Sutter, M. Tomamichel, and A. W. Harrow. Monotonicity of relative entropy via pinched Petz recovery map, 2015. [arXiv:1507.00303](#).
- [44] A. Uhlmann. Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory. *Communications in Mathematical Physics*, 54(1):21–32, 1977.
- [45] M. M. Wilde. Recoverability in quantum information theory. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 471(2182):20150338, 2015.
- [46] A. Winter. Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints, 2015. [arXiv:1507.07775](#).
- [47] A. Winter and K. Li. A stronger subadditivity relation? With applications to squashed entanglement, sharability and separability, 2012. http://www.maths.bris.ac.uk/~csajw/stronger_subadditivity.pdf.